Correlation tests for high-dimensional data using extended cross-data-matrix methodology

著者	Yata Kazuyoshi, Aoshima Makoto
journal or publication title	Journal of multivariate analysis
volume	117
page range	313-331
year	2013-05
権利	(C) 2013 Elsevier Inc. NOTICE: this is the author's version of a work that was accepted for publication in Journal of Multivariate Analysis. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Journal of Multivariate Analysis, Vol.117, Pages:313-331. doi:10.1016/j.jmva.2013.03.007.
URL	http://hdl.handle.net/2241/118867

doi: 10.1016/j.jmva.2013.03.007

Correlation tests for high-dimensional data using extended cross-data-matrix methodology

Kazuyoshi Yata[a], Makoto Aoshima[a,1,]

[a]*Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan*

Abstract

In this paper, we consider tests of correlation when the sample size is much lower than the dimension. We propose a new estimation methodology called the *extended cross-data-matrix methodology*. By applying the method, we give a new test statistic for high-dimensional correlations. We show that the test statistic is asymptotically normal when $p \to \infty$ and $n \to \infty$. We propose a test procedure along with sample size determination to assure both prespecified size and power for testing high-dimensional correlations. We further develop a multiple testing procedure to control both family wise error rate and power. Finally, we demonstrate how the test procedures perform in actual data analyses by using two microarray data sets.

Keywords: Cross-data-matrix methodology; Graphical modeling; HDLSS; High-dimensional regression; Pathway analysis; Two-stage procedure.

1. Introduction

A common feature of high-dimensional data is that the data dimension is high, however, the sample size is relatively small. This is the so-called "HDLSS" or "large p, small n" data situation where $p/n \to \infty$; here p is the data dimension and n is the sample size. The asymptotic studies of this type of data are becoming increasingly relevant. In recent years, substantial work had been done on the asymptotic behavior of eigenvalues of the sample covariance matrix in the limit as $p \to \infty$, see Johnstone [19]

Email address: aoshima@math.tsukuba.ac.jp (Makoto Aoshima)

[1]Phone & Fax: +81-298-53-6501

and Paul [22] for Gaussian assumptions and Baik and Silverstein [6] for non-Gaussian but i.i.d. assumptions. Those literatures handled the cases when p and n increase at the same rate, i.e. $p/n \to c > 0$. The asymptotic behaviors of high-dimensional, low-sample-size (HDLSS) data were studied by Ahn et al. [1], Hall et al. [16], and Yata and Aoshima [31] when $p \to \infty$ while n is fixed. They explored conditions to give a geometric representation of HDLSS data. The HDLSS asymptotic study usually assumes either the normality as the population distribution or a ρ-mixing condition as the dependency of random variables in a sphered data matrix. See Jung and Marron [20]. Yata and Aoshima [29] succeeded in investigating the consistency properties of both eigenvalues and eigenvectors of the sample covariance matrix in more general settings including the case that all eigenvalues are in the range of sphericity. In addition, Yata and Aoshima [30] created the *cross-data-matrix (CDM) methodology* that provides effective inference on the eigenspace for HDLSS data. Recently, Chen and Qin [9] gave a two-sample test for high-dimensional data. Aoshima and Yata [2, 3] developed a variety of high-dimensional statistical inference based on the geometric representations and gave sample size determination to assure prespecified accuracy. In this paper, we consider tests of correlation coefficients for high-dimensional data and give sample size determination to assure prespecified accuracy.

Let $\boldsymbol{x}_{1(*)}, \boldsymbol{x}_{2(*)}, ...$ be a sequence of i.i.d. $p+1$-variate data vectors, where $\boldsymbol{x}_{j(*)} = (\boldsymbol{x}_j^T, x_{j(*)})^T$ with $\boldsymbol{x}_j = (x_{1j}, ..., x_{pj})^T$. Having recorded $\boldsymbol{x}_{1(*)}, ..., \boldsymbol{x}_{n(*)}$. We assume $n \geq 4$. Here, $\boldsymbol{x}_j$ has unknown mean vector, $\boldsymbol{\mu}$, and unknown covariance matrix, $\boldsymbol{\Sigma}$ ($\geq \boldsymbol{O}$), and $x_{j(*)}$ has unknown mean vector, μ_*, and unknown variance, $\sigma_*^2 \in (0, \infty)$. Let $\boldsymbol{\theta} = (\mu_*, \sigma_*^2, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. We denote the covariance vector between $\boldsymbol{x}_j$ and $x_{j(*)}$ by $\mathrm{Cov}_{\boldsymbol{\theta}}(\boldsymbol{x}_j, x_{j(*)}) = \boldsymbol{\sigma}$. We denote the correlation coefficient vector between $\boldsymbol{x}_j$ and $x_{j(*)}$ by $\mathrm{Corr}_{\boldsymbol{\theta}}(\boldsymbol{x}_j, x_{j(*)}) = \boldsymbol{\rho}$. We consider testing the correlation between $\boldsymbol{x}_j$ and $x_{j(*)}$ by

$$H_0 : \boldsymbol{\rho} = \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\rho} \neq \mathbf{0}. \tag{1}$$

The test of the correlation is a very important tool of pathway analysis or graphical modeling for high-dimensional data. For example, Drton and Perlman [11] and Wille et al. [27] considered pathway analysis or graphical modeling of microarray data by testing an individual correlation coefficient. On the other hand, Hero and Rajaratnam [17] considered correlation screening procedures for high-dimensional data by using a test of correlations. Zhong and Chen [32] considered tests of a regression coefficient vector on linear regression models. Aoshima and Yata [2] created a test statistic for (1) by using the CDM methodology.

Let $\boldsymbol{\Sigma} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^T$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues, $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$, and $\boldsymbol{H}$ is an orthogonal matrix of corresponding eigenvectors. We assume that $\limsup_{p\to\infty} \mathrm{tr}(\boldsymbol{\Sigma})/p < \infty$. Let $\boldsymbol{x}_j = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}\boldsymbol{z}_j + \boldsymbol{\mu}$, $j = 1, ..., n$. Then, $E(\boldsymbol{z}_j) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{z}_j) = \boldsymbol{I}_p$, where $\boldsymbol{I}_p$ denotes the identity matrix of dimension p. In this paper, we assume the following model:

$$\boldsymbol{x}_j = \boldsymbol{\Gamma}\boldsymbol{w}_j + \boldsymbol{\mu}, \tag{2}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_r)$ is a $p \times r$ matrix for some $r > 0$ such that $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \boldsymbol{\Sigma}$, and $\boldsymbol{w}_j$, $j = 1, ..., n$, are i.i.d. random vectors having $E(\boldsymbol{w}_j) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{w}_j) = \boldsymbol{I}_r$. See also Bai and Saranadasa [5] and Chen and Qin [9]. Note that the model (2) includes the case that $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$ and $\boldsymbol{w}_j = \boldsymbol{z}_j$. As for $\boldsymbol{w}_j = (w_{1j}, ..., w_{rj})^T$, we assume that

(A-i) The fourth moments of w_{ij}s are uniformly bounded, and w_{ij}, $i = 1, ..., r$, are independent.

We assume the following assumption for $\boldsymbol{\Sigma}$ as necessary:

(A-ii) $\dfrac{\mathrm{tr}(\boldsymbol{\Sigma}^4)}{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2} \to 0$ as $p \to \infty$.

Remark 1. If all λ_is are bounded, (A-ii) is trivially true. For a spiked model such as $\lambda_i = a_i p^{\alpha_i}$ $(i = 1, ..., m)$ and $\lambda_i = c_i$ $(i = m+1, ..., p)$ with positive constants a_is, c_is and α_is, (A-ii) is true under the conditions that $\alpha_i < 1/2, i = 1, ..., m$ and $m < \infty$. See Yata and Aoshima [30] for the details of a spiked model. For $\boldsymbol{\Sigma} = c(\rho^{|i-j|^q})$ with c (> 0), q (> 0) and $\rho \in (0, 1)$, (A-ii) holds. In addition, for the above cases, it holds that $\mathrm{tr}(\boldsymbol{\Sigma}^2) = O(p)$.

Let

$$x_{j(*)} = c_* w_{j*} + \sum_{i=1}^{r} c_i w_{ij} + \mu_*, \tag{3}$$

where c_* and c_is are constants such that $c_*^2 + \sum_{i=1}^r c_i^2 = \sigma_*^2$, and w_{j*} is a random variable such that $E(w_{j*}) = 0$, $E(w_{j*}^2) = 1$, and $E(w_{ij}w_{j*}) = 0$ for $i = 1, ..., r$. Note that $\sum_{i=1}^r c_i\boldsymbol{\gamma}_i = \boldsymbol{\sigma}$ and $\sum_{i,i'}^r c_i c_{i'} \boldsymbol{\gamma}_i^T \boldsymbol{\gamma}_{i'} = ||\boldsymbol{\sigma}||^2$, where $||\cdot||$ denotes the Euclidean norm. Then, $\boldsymbol{x}_j$ and $x_{j(*)}$ $(j = 1, ..., n)$ are uncorrelated when $\sum_{i=1}^r c_i\boldsymbol{\gamma}_i = \boldsymbol{0}$. In this paper, we assume the following assumption for w_{j*} as necessary:

(A-iii) The fourth moment of w_{j*} is bounded, and w_{j*} and $\boldsymbol{w}_j$ are independent.

If $\boldsymbol{x}_{j(*)}$ is Gaussian, (A-i) and (A-iii) hold.

Remark 2. Let $w_{r+1j} = w_{j*}$ for each j. We consider the following assumption: The fourth moments of w_{ij}, $i = 1, ..., r+1$, are bounded, and $E(w_{l_1 j}^{\alpha_1} w_{l_2 j}^{\alpha_2} \cdots w_{l_q j}^{\alpha_q}) = E(w_{l_1 j}^{\alpha_1}) E(w_{l_2 j}^{\alpha_2}) \cdots E(w_{l_q j}^{\alpha_q})$ for all $l_1 \neq l_2 \neq \cdots \neq l_q \in [1, r+1]$, where α_is are integers within $[0, 4]$ such that $\sum_{i=1}^{q} \alpha_i \leq 8$. See Chen and Qin [9] and Zhong and Chen [32] for the assumption. Then, we can claim all the results in this paper under the assumption instead of (A-i) and (A-iii).

Throughout this paper, we write that

$$\boldsymbol{S}_n = \sum_{j=1}^{n} \frac{(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_n)(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_n)^T}{n-1}, \quad S_{n(*)} = \sum_{j=1}^{n} \frac{(x_{j(*)} - \overline{x}_{n(*)})^2}{n-1}$$
$$\text{and} \quad \boldsymbol{s}_{n(*)} = \sum_{j=1}^{n} \frac{(x_{j(*)} - \overline{x}_{n(*)})(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_n)}{n-1}, \tag{4}$$

where $\overline{\boldsymbol{x}}_n = n^{-1} \sum_{j=1}^{n} \boldsymbol{x}_j$ and $\overline{x}_{n(*)} = n^{-1} \sum_{j=1}^{n} x_{j(*)}$. When $n > p$, one may consider a multiple correlation coefficient by $\rho = (\boldsymbol{\sigma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\sigma} / \sigma_*^2)^{1/2}$. Then, a test statistic of (1) is given by $\widehat{\rho} = (\boldsymbol{s}_{n(*)}^T \boldsymbol{S}_n^{-1} \boldsymbol{s}_{n(*)} / S_{n(*)})^{1/2}$. When $\boldsymbol{x}_{j(*)}$ is Gaussian, a certain transformation of $\widehat{\rho}$ is distributed as an F-distribution. See, for example, Chapter 4 in Fujikoshi et al. [13]. However, in the HDLSS context where $p > n$, $\widehat{\rho}$ does not work since the inverse matrix of $\boldsymbol{S}_n$ does not exist. Several authors considered substituting some estimators such as the Moore-Penrose inverse matrix for $\boldsymbol{S}_n^{-1}$. See Srivastave [25] for example. Yata and Aoshima [31] applied a method called the *noise-reduction methodology* to estimating $\boldsymbol{\Sigma}^{-1}$ and compared performance of estimators of $\boldsymbol{\Sigma}^{-1}$. Refer to Sections 7 and 8 of Yata and Aoshima [31]. As for a test of independence for high-dimensional data, one may refer to Székely et al. [26] about distance correlation.

In this paper, we provide test procedures for correlations appeared in HDLSS data. In Section 2, we propose a new estimation method called the *extended cross-data-matrix methodology.* By applying the method, we give a new test statistic for high-dimensional correlations. We show that the test statistic is asymptotically normal when $p \to \infty$ and $n \to \infty$. In Section 3, we propose a test procedure along with sample size determination to assure both prespecified size and power for testing high-dimensional correlations. In Section 4, we develop a multiple testing procedure to control both family wise error rate and power. Finally, in Section 5, we demonstrate how the test procedures perform in actual data analyses by using two microarray data sets.

2. Test of high-dimensional correlations

Throughout this paper, we consider applying the following new estimation method called the *extended cross-data-matrix (ECDM) methodology*. The ECDM methodology is considered as an extension of the CDM methodology developed by Yata and Aoshima [30].

2.1. Extended cross-data-matrix (ECDM) methodology

Let $n_{(1)} = \lceil n/2 \rceil$ and $n_{(2)} = n - n_{(1)}$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. Now, we consider two sets $\boldsymbol{V}_{n(1)(k)}$ and $\boldsymbol{V}_{n(2)(k)}$ $(k = 3, ..., 2n-1)$ such that $\#(\boldsymbol{V}_{n(l)(k)}) = n_{(l)}$, $l = 1, 2$, $\boldsymbol{V}_{n(1)(k)} \cap \boldsymbol{V}_{n(2)(k)} = \boldsymbol{\emptyset}$, $\boldsymbol{V}_{n(1)(k)} \cup \boldsymbol{V}_{n(2)(k)} = \{1, ..., n\}$ and

$$i \in \boldsymbol{V}_{n(1)(i+j)} \quad \text{and} \quad j \in \boldsymbol{V}_{n(2)(i+j)} \quad \text{for} \quad i < j \ (\leq n), \tag{5}$$

where $\#(\boldsymbol{S})$ denotes the number of elements in a set $\boldsymbol{S}$. Then, we find the two sets as follows:

$$\boldsymbol{V}_{n(1)(k)} = \begin{cases} \big\{ \lfloor k/2 \rfloor - n_{(1)} + 1, ..., \lfloor k/2 \rfloor \big\} & \text{if} \ \ \lfloor k/2 \rfloor \geq n_{(1)}, \\ \big\{ 1, ..., \lfloor k/2 \rfloor \big\} \cup \big\{ \lfloor k/2 \rfloor + n_{(2)} + 1, ..., n \big\} & \text{otherwise} \end{cases}$$

and

$$\boldsymbol{V}_{n(2)(k)} = \begin{cases} \big\{ \lfloor k/2 \rfloor + 1, ..., \lfloor k/2 \rfloor + n_{(2)} \big\} & \text{if} \ \ \lfloor k/2 \rfloor \leq n_{(1)}, \\ \big\{ 1, ..., \lfloor k/2 \rfloor - n_{(1)} \big\} \cup \big\{ \lfloor k/2 \rfloor + 1, ..., n \big\} & \text{otherwise} \end{cases}$$

for $k = 3, ..., 2n-1$, where $\lfloor x \rfloor$ denotes the largest integer $\leq x$. The ECDM methodology is a method to provide an unbiased estimator by using $\boldsymbol{V}_{n(1)(i+j)}$ and $\boldsymbol{V}_{n(2)(i+j)}$.

Let

$$\overline{\boldsymbol{x}}_{n(1)(k)} = n_{(1)}^{-1} \sum_{j \in \boldsymbol{V}_{n(1)(k)}} \boldsymbol{x}_j, \qquad \overline{\boldsymbol{x}}_{n(2)(k)} = n_{(2)}^{-1} \sum_{j \in \boldsymbol{V}_{n(2)(k)}} \boldsymbol{x}_j,$$

$$\overline{x}_{n(1*)(k)} = n_{(1)}^{-1} \sum_{j \in \boldsymbol{V}_{n(1)(k)}} x_{j(*)}, \quad \text{and} \quad \overline{x}_{n(2*)(k)} = n_{(2)}^{-1} \sum_{j \in \boldsymbol{V}_{n(2)(k)}} x_{j(*)}$$

for $k = 3, ..., 2n-1$. Note that $E_{\boldsymbol{\theta}}\{(x_{i(*)} - \overline{x}_{n(1*)(i+j)})(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n(1)(i+j)})\} = n_{(1)}^{-1}(n_{(1)}-1)\boldsymbol{\sigma}$ and $E_{\boldsymbol{\theta}}\{(x_{j(*)} - \overline{x}_{n(2*)(i+j)})(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(i+j)})\} = n_{(2)}^{-1}(n_{(2)}-1)\boldsymbol{\sigma}$ for $i < j (\leq n)$. From (5), we emphasize the following facts:

(i) $\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n(1)(i+j)}$ and $\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(i+j)}$ are independent;
(ii) $x_{i(*)} - \overline{x}_{n(1*)(i+j)}$ and $x_{j(*)} - \overline{x}_{n(2*)(i+j)}$ are independent

for $i < j$ $(\leq n)$. We propose an estimator of $||\boldsymbol{\sigma}||^2$ by

$$\begin{aligned} \widehat{T}_{n,\boldsymbol{\sigma}} = \frac{2u_n}{n(n-1)} \sum_{i<j}^{n} &(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n(1)(i+j)})^T (\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(i+j)}) \\ &\times (x_{i(*)} - \overline{x}_{n(1*)(i+j)})(x_{j(*)} - \overline{x}_{n(2*)(i+j)}), \end{aligned} \tag{6}$$

where $u_n = n_{(1)}n_{(2)}/\{(n_{(1)}-1)(n_{(2)}-1)\}$. Then, we note that $E_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}}) = ||\boldsymbol{\sigma}||^2$. Let $\mathrm{Var}(w_{ij}^2) = M_i$, $i = 1, ..., r$. Let $\sigma_*^4 = (\sigma_*^2)^2$. Under (A-i) to (A-iii), from Lemma A.1 in Appendix, we have as $p \to \infty$ and $n \to \infty$ that

$$\begin{aligned} \mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}}) = \Big\{ \frac{2\sigma_*^4 \mathrm{tr}(\boldsymbol{\Sigma}^2)}{n^2} + \frac{4}{n}\Big(&\sigma_*^2 \boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma} + ||\boldsymbol{\sigma}||^4 \\ &+ \sum_{i=1}^{r} (M_i - 2) c_i^2 (\boldsymbol{\sigma}^T \boldsymbol{\gamma}_i)^2 \Big) \Big\} \{1 + o(1)\}. \end{aligned} \tag{7}$$

Remark 3. Another unbiased estimator of $||\boldsymbol{\sigma}||^2$ is

$$\begin{aligned} \widehat{T}_{n,\boldsymbol{\sigma}(AY)} = \frac{1}{(n_{(1)}-1)(n_{(2)}-1)} \sum_{i=1}^{n_{(1)}} \sum_{j=n_{(1)}+1}^{n} &(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n(1)(n+1)})^T (\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(n+1)}) \\ &\times (x_{i(*)} - \overline{x}_{n(1*)(n+1)})(x_{j(*)} - \overline{x}_{n(2*)(n+1)}) \end{aligned}$$

that was given by Aoshima and Yata [2] when applying the CDM methodology. Then it holds as $p \to \infty$ and $n \to \infty$ that $\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}(AY)}) = [4\sigma_*^4 \mathrm{tr}(\boldsymbol{\Sigma}^2)$

$/n^2+4\{\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma}+||\boldsymbol{\sigma}||^4+\sum_{i=1}^r(M_i-2)c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2\}/n]\{1+o(1)\}$ under (A-i) to (A-iii). Thus the asymptotic variance of $\widehat{T}_{n,\boldsymbol{\sigma}}$ is smaller than that of $\widehat{T}_{n,\boldsymbol{\sigma}(AY)}$. The ECDM methodology is not a resampling-based extension of the CDM methodology. The ECDM methodology considers the combination of cross data matrices so as to construct an unbiased estimator efficiently and enjoy desirable properties in non-Gaussian situations. See Section 2.5 for the details. As for a resampling-based extension, see Aoshima and Yata [4].

Remark 4. One can save the computational cost of $\widehat{T}_{n,\boldsymbol{\sigma}}$ by substituting previously calculated $\overline{\boldsymbol{x}}_{n(i)(k)}$s and $\overline{x}_{n(i*)(k)}$s in (6). Then, the computational cost of $\widehat{T}_{n,\boldsymbol{\sigma}}$ is written by the order, $O(n^2p)$.

2.2. Asymptotic distribution of $\widehat{T}_{n,\boldsymbol{\sigma}}$

We assume the following extra assumption:

(A-iv) $\quad \liminf \dfrac{\sigma_*^2\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}}{n||\boldsymbol{\sigma}||^2} > 0$ as $p\to\infty$ and $n\to\infty$ when $||\boldsymbol{\sigma}||^2\neq 0$.

Then, we have the following theorem.

Theorem 2.1. *Assume (A-i) to (A-iv). It holds as* $p\to\infty$ *and* $n\to\infty$ *that*

$$\frac{\widehat{T}_{n,\boldsymbol{\sigma}}-||\boldsymbol{\sigma}||^2}{\sqrt{Var_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})}}\Rightarrow N(0,1), \tag{8}$$

where "$\Rightarrow$" denotes the convergence in distribution and $N(0,1)$ *denotes a random variable distributed as the standard normal distribution.*

If one cannot assume (A-iv), we have the following result.

Corollary 2.1. *Assume (A-i) and (A-iii). Assume* $\sigma_*^2\sqrt{tr(\boldsymbol{\Sigma}^2)}/(n||\boldsymbol{\sigma}||^2)=o(1)$. *Then, it holds as* $p\to\infty$ *and* $n\to\infty$ *that*

$$\frac{\widehat{T}_{n,\boldsymbol{\sigma}}}{||\boldsymbol{\sigma}||^2}=1+o_p(1).$$

We emphasize that the assertion in Theorem 2.1 is still claimed under the HDLSS setting where $p/n \to \infty$. From the facts that $\sum_{i=1}^{r} c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2 \le \sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma} \le \sigma_*^2||\boldsymbol{\sigma}||^2\lambda_1 \le \sigma_*^2||\boldsymbol{\sigma}||^2\mathrm{tr}(\boldsymbol{\Sigma}^4)^{1/4}$ and $\limsup n||\boldsymbol{\sigma}||^2/\{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma}^2)^{1/2}\} < \infty$ under (A-iv), we note that $\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})/\{2\sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^2)/n^2\} = 1 + o(1)$ as $p \to \infty$ and $n \to \infty$ under (A-i) to (A-iv). Since $\boldsymbol{\Sigma}$ and σ_*^2 are unknown, it is necessary to estimate $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ and σ_*^2. By applying the ECDM methodology, we propose an estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ by

$$W_n = \frac{2u_n}{n(n-1)}\sum_{i<j}^{n}\{(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n(1)(i+j)})^T(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(i+j)})\}^2. \tag{9}$$

We note that $E_{\boldsymbol{\theta}}(W_n) = \mathrm{tr}(\boldsymbol{\Sigma}^2)$. As for the variance of W_n, see Section 2.5. Then, we have the following result.

Corollary 2.2. *Assume (A-i) to (A-iv). It holds as $p \to \infty$ and $n \to \infty$ that*

$$\frac{\widehat{T}_{n,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2}{S_{n(*)}\sqrt{2W_n}/n} \Rightarrow N(0,1).$$

Remark 5. From (7), under (A-i) to (A-iv), it holds as $p \to \infty$ and $n \to \infty$ that

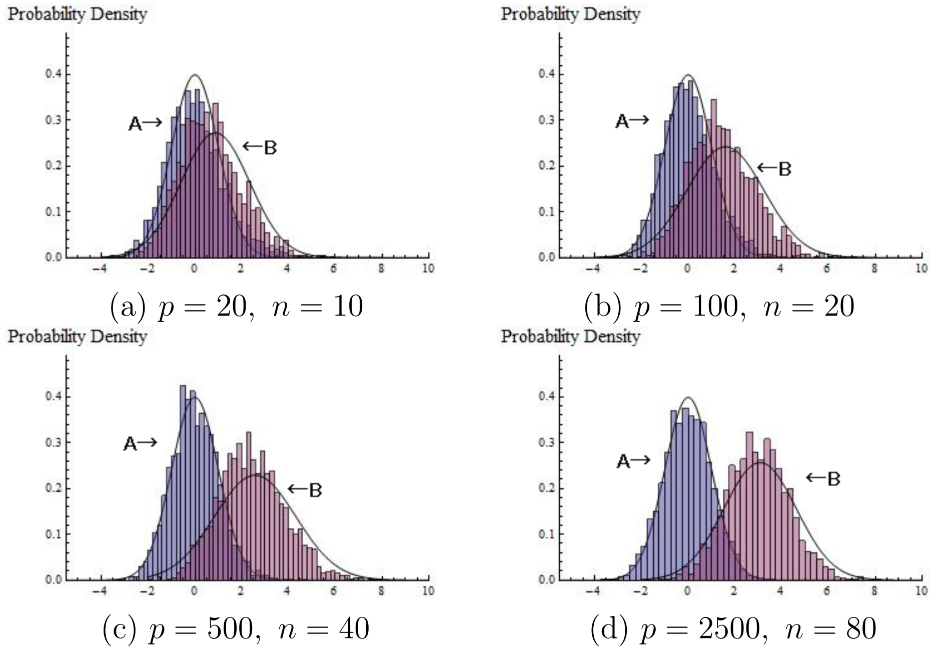
$$\begin{aligned}&\frac{\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})}{2\sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^2)/n^2}\\ &= 1 + \frac{2n\big\{\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma} + ||\boldsymbol{\sigma}||^4 + \sum_{i=1}^{r}(M_i - 2)c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2\big\}}{\sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^2)} + o(1) = 1 + o(1).\end{aligned}$$

Thus one may write (8) as

$$\frac{\widehat{T}_{n,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2}{\sigma_*^2\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)(1+u)/n}} \Rightarrow N(0,1),$$

where $u = 2n\{\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma} + ||\boldsymbol{\sigma}||^4 + \sum_{i=1}^{r}(M_i - 2)c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2\}/\{\sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^2)\}$.

Let us observe Corollary 2.2 in view of Remark 5. Now, we considered an easy example such as $\boldsymbol{\mu} = \boldsymbol{0}$, $\boldsymbol{\Sigma} = (0.3^{|i-j|^{1/3}})$, $\mu_* = 0$, $\sigma_*^2 = 1$, $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$ and $\boldsymbol{\rho} = \boldsymbol{0}$ ($c_1 = \cdots = c_r = 0$, $c_* = 1$) or $\boldsymbol{\rho} \ne \boldsymbol{0}$ ($||\boldsymbol{\sigma}||^2 = \sum_{i=1}^{10}\lambda_i/20$; $c_1 = \cdots = c_{10} = \sqrt{1/20}$, $c_* = \sqrt{1/2}$ and the other c_is are 0). Note that



(a) $p = 20,\ n = 10$ (b) $p = 100,\ n = 20$

(c) $p = 500,\ n = 40$ (d) $p = 2500,\ n = 80$

Figure 1: The solid lines are probability densities of $A : N(0, 1)$ and $B : N(\omega, 1 + u)$. The histograms of $\widehat{T}_{n,\boldsymbol{\sigma}}/\{S_{n(*)}\sqrt{2W_n}/n\}$ for both the cases of $\boldsymbol{\rho} = \mathbf{0}$ and $\boldsymbol{\rho} \neq \mathbf{0}$ fit the solid lines with increasing dimension and sample size: (a) $(p, n) = (20, 10)$, (b) $(p, n) = (100, 20)$, (c) $(p, n) = (500, 40)$, and (d) $(p, n) = (2500, 80)$.

$r = p$ and $||\boldsymbol{\sigma}||^2 = \sum_{i=1}^{p} c_i^2\lambda_i$ from $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$. We considered four cases: (a) $p = 20,\ n = 10$, (b) $p = 100,\ n = 20$, (c) $p = 500,\ n = 40$, and (d) $p = 2500,\ n = 80$. Fig.1 (a), (b), (c) and (d) give the histograms of 2000 independent outcomes of $\widehat{T}_{n,\boldsymbol{\sigma}}/\{S_{n(*)}\sqrt{2W_n}/n\}$ both when $\boldsymbol{\rho} = \mathbf{0}$ and $\boldsymbol{\rho} \neq \mathbf{0}$. Here, $\boldsymbol{x}_j,\ j = 1, ..., n$, were generated independently from a pseudorandom normal distribution with mean vector zero and covariance matrix $\boldsymbol{\Sigma}$ for each case of $(p, n) = (20, 10),\ (100, 20),\ (500, 40)$ and $(2500, 80)$. Note that $M_i = 2,\ i = 1, ..., p$. Independent of $\boldsymbol{x}_j,\ w_{j*},\ j = 1, ..., n$, were generated independently from a pseudorandom standard normal distribution. Let $\omega = ||\boldsymbol{\sigma}||^2/(\sigma_*^2\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}/n)$. From Corollary 2.2 in view of Remark 5, we expected that $\widehat{T}_{n,\boldsymbol{\sigma}}/(S_{n(*)}\sqrt{2W_n}/n)$ is close to $N(0, 1)$ when $\boldsymbol{\rho} = \mathbf{0}$ and $\widehat{T}_{n,\boldsymbol{\sigma}}/(S_{n(*)}\sqrt{2W_n}/n)$ is close to $N(\omega, 1 + u)$ when $\boldsymbol{\rho} \neq \mathbf{0}$. When $p = 20$ and $p = 100$, the histograms appear different from the probability densities especially when $\boldsymbol{\rho} \neq \mathbf{0}$. However, as expected, the histograms fit well the probability densities as p and n increase.

2.3. Test of correlations

We are interested in designing a test of (1) having size α, where $\alpha \in (0, 1/2)$ is a prespecified constant. We test the hypothesis (1) by

$$\text{rejecting } H_0 \Longleftrightarrow \frac{\widehat{T}_{n,\boldsymbol{\sigma}}}{S_{n(*)}\sqrt{2W_n/n}} > z_\alpha, \tag{10}$$

where z_α is a constant such that $P\{N(0,1) > z_\alpha\} = \alpha$. Then, we have the following theorem.

Theorem 2.2. *Under (A-i) to (A-iv), the test by (10) has that*

$$size = \alpha + o(1) \quad and \quad power = \Phi\Big(\frac{n||\boldsymbol{\sigma}||^2}{\sigma_*^2\sqrt{2tr(\boldsymbol{\Sigma}^2)}} - z_\alpha\Big) + o(1) \tag{11}$$

as $p \to \infty$ and $n \to \infty$, where $\Phi(\cdot)$ denotes the c.d.f. of $N(0,1)$.

When (A-iv) is not met, we have the following result.

Corollary 2.3. *Assume (A-i) to (A-iii). Assume $\sigma_*^2\sqrt{tr(\boldsymbol{\Sigma}^2)}/(n||\boldsymbol{\sigma}||^2) = o(1)$ when $\boldsymbol{\rho} \neq \mathbf{0}$. Then, the test by (10) has that*

$$size = \alpha + o(1) \quad and \quad power = 1 + o(1)$$

as $p \to \infty$ and $n \to \infty$.

Remark 6. From Remark 5, one may write the power in (11) as

$$power = \Phi\Big(\frac{n||\boldsymbol{\sigma}||^2}{\sigma_*^2\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)(1+u)}} - \frac{z_\alpha}{\sqrt{1+u}}\Big) + o(1),$$

where $u = 2n\{\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma} + ||\boldsymbol{\sigma}||^4 + \sum_{i=1}^r (M_i - 2)c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2\}/\{\sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^2)\}$.

2.4. Moderate sample performances

In order to study the performance of the test by (10), we used computer simulations. We set $\alpha = 0.05$. We generated $\boldsymbol{x}_j$s independently from a pseudorandom normal distribution with mean vector zero and covariance matrix $\boldsymbol{\Sigma}$. Independent of $\boldsymbol{x}_j$s, we generated w_{j*}s independently from a pseudorandom normal distribution with zero mean and unit variance. We considered $\sigma_*^2 = 1$, $\boldsymbol{\Sigma} = (0.3^{|i-j|^{1/3}})$ and $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$. Note that $r = p$

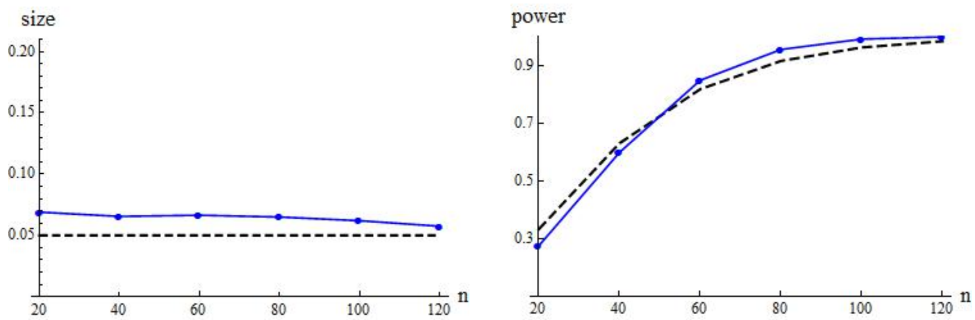


Figure 2: When $p = 1000$ and $n = 20(20)120$, the size $\alpha = 0.05$ (broken line) and $\overline{\alpha}$ (solid line) are displayed in the left panel. The power $\Phi[\{2\sigma_*^4 \mathrm{tr}(\boldsymbol{\Sigma}^2)(1+u)\}^{-1/2} n||\boldsymbol{\sigma}||^2 - (1+u)^{-1/2} z_\alpha]$ (broken line) and $1 - \overline{\beta}$ (solid line) are displayed in the right panel.

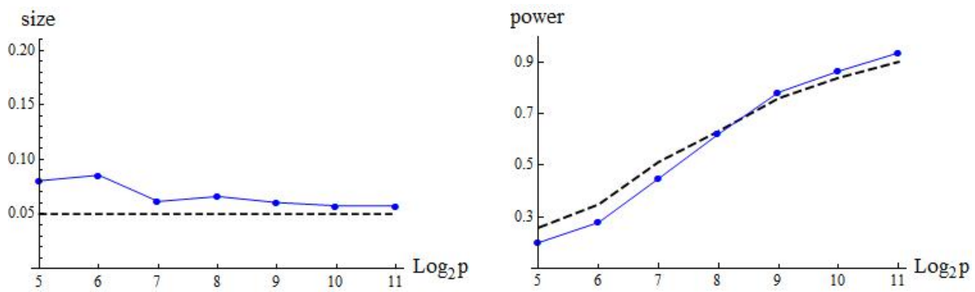


Figure 3: When $p = 2^s$ $(s = 5, ..., 11)$ and $n = 2\lceil p^{1/2} \rceil$, the size $\alpha = 0.05$ (broken line) and $\overline{\alpha}$ (solid line) are displayed in the left panel. The power $\Phi[\{2\sigma_*^4 \mathrm{tr}(\boldsymbol{\Sigma}^2)(1+u)\}^{-1/2} n||\boldsymbol{\sigma}||^2 - (1+u)^{-1/2} z_\alpha]$ (broken line) and $1 - \overline{\beta}$ (solid line) are displayed in the right panel.

and $||\boldsymbol{\sigma}||^2 = \sum_{i=1}^{p} c_i^2 \lambda_i$ from $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$. We considered two cases: (i) $\boldsymbol{\rho} = \mathbf{0}$ $(c_1 = \cdots = c_p = 0,\ c_* = 1)$ and (ii) $\boldsymbol{\rho} \neq \mathbf{0}$ $(||\boldsymbol{\sigma}||^2 = \lambda_5/2;\ c_5 = \sqrt{1/2},\ c_* = \sqrt{1/2}$ and the other c_is are 0). In Fig. 2, we set $p = 1000$ and $n = 20(20)120$. In Fig. 3, we set $p = 2^s$ $(s = 5, ..., 11)$ and $n = 2\lceil p^{1/2} \rceil$. The findings were obtained by averaging the outcomes from 4000 $(= R$, say) replications, where the first 2000 replications were generated for (i), and the last 2000 replications were generated for (ii).

Under a fixed scenario, suppose that the rth replication ends with a test result given by (10). We defined $P_r = 1$ (or 0) accordingly as $H_0 :\ \boldsymbol{\rho} = \mathbf{0}$ was falsely rejected (or not) and $H_1 :\ \boldsymbol{\rho} \neq \mathbf{0}$ was falsely rejected (or not). We defined $\overline{\alpha} = (R/2)^{-1} \sum_{r=1}^{R/2} P_r$ to estimate the size and $1 - \overline{\beta} = 1 - (R/2)^{-1} \sum_{r=R/2+1}^{R} P_r$ to estimate the power. Note that the standard error of the simulation study was no more than 0.0112. Throughout, we observed that the test by (10) showed good performances as described in Theorem 2.2 (or Remark 6) as p and n increase.

2.5. Comparison of estimators for tr($\boldsymbol{\Sigma}^2$)

From Lemma A.2 in Appendix, we have as $p \to \infty$ and $n \to \infty$ that

$$\mathrm{Var}_{\boldsymbol{\theta}}\Big(\frac{W_n}{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\Big) = \Big(\frac{4}{n^2} + \frac{8\mathrm{tr}(\boldsymbol{\Sigma}^4) + 4\sum_{i=1}^r (M_i - 2)(\boldsymbol{\gamma}_i^T \boldsymbol{\Sigma}\boldsymbol{\gamma}_i)^2}{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 n}\Big)\{1 + o(1)\}$$

under (A-i). Further, if $\boldsymbol{x}_j$ is Gaussian, we have as $p \to \infty$ and $n \to \infty$ that $\mathrm{Var}_{\boldsymbol{\theta}}\{W_n/\mathrm{tr}(\boldsymbol{\Sigma}^2)\} = [4/n^2 + 8\mathrm{tr}(\boldsymbol{\Sigma}^4)/\{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 n\}]\{1 + o(1)\}$. Yata [28] applied the CDM methodology due to Yata and Aoshima [30] to obtaining an unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ by $\mathrm{tr}(\boldsymbol{S}_{n(1)}\boldsymbol{S}_{n(2)})$, where $\boldsymbol{S}_{n(1)} = (n_{(1)} - 1)^{-1}\sum_{j=1}^{n_{(1)}}(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(1)(n+1)})(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(1)(n+1)})^T$ and $\boldsymbol{S}_{n(2)} = (n_{(2)} - 1)^{-1}\sum_{j=n_{(1)}+1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(n+1)})(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(n+1)})^T$. Note that $E_{\boldsymbol{\theta}}\{\mathrm{tr}(\boldsymbol{S}_{n(1)}\boldsymbol{S}_{n(2)})\} = \mathrm{tr}(\boldsymbol{\Sigma}^2)$. Then, it holds as $p \to \infty$ and $n \to \infty$ that

$$\begin{aligned}&\mathrm{Var}_{\boldsymbol{\theta}}\Big(\frac{\mathrm{tr}(\boldsymbol{S}_{n(1)}\boldsymbol{S}_{n(2)})}{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\Big)\\ &= \Big(\frac{8}{n^2} + \frac{8\mathrm{tr}(\boldsymbol{\Sigma}^4) + 4\sum_{i=1}^r (M_i - 2)(\boldsymbol{\gamma}_i^T \boldsymbol{\Sigma}\boldsymbol{\gamma}_i)^2}{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 n}\Big)\{1 + o(1)\}\end{aligned}$$

under (A-i). Thus the asymptotic variance of W_n is smaller than that of $\mathrm{tr}(\boldsymbol{S}_{n(1)}\boldsymbol{S}_{n(2)})$. On the other hand, Bai and Saranadasa [5] and Srivastava [24] considered an estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ by $W_{n(BS)} = c_n^{-1}\{\mathrm{tr}(\boldsymbol{S}_n^2) - \mathrm{tr}(\boldsymbol{S}_n)^2/(n-1)\}$ with $c_n = (n-2)(n+1)/(n-1)^2$ under the Gaussian assumption. They showed that, when $\boldsymbol{x}_j$ is Gaussian, it holds that $E_{\boldsymbol{\theta}}(W_{n(BS)}) = \mathrm{tr}(\boldsymbol{\Sigma}^2)$ and

$$\mathrm{Var}_{\boldsymbol{\theta}}\Big(\frac{W_{n(BS)}}{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\Big) = \Big(\frac{4}{n^2} + \frac{8\mathrm{tr}(\boldsymbol{\Sigma}^4)}{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 n}\Big)\{1 + o(1)\}.$$

Thus the ECDM methodology is desirable in the sense that the asymptotic variance of W_n is equivalent to that of $W_{n(BS)}$ which specializes the Gaussian case. It should be noted that $W_{n(BS)}$ is biased unless $\boldsymbol{x}_j$ is Gaussian. In addition, one cannot claim $\mathrm{Var}_{\boldsymbol{\theta}}\{W_{n(BS)}/tr(\boldsymbol{\Sigma}^2)\} < \infty$ unless the eighth moments of each variable in $\boldsymbol{w}_j$ are uniformly bounded. Contrary to that, the proposed estimator, W_n, is robust in non-Gaussian situations.

On the other hand, Zhong and Chen [32] considered an unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ by $W_{n(Z)} = \{n(n-1)\}^{-1}\sum_{i\neq j}^{n}(\boldsymbol{x}_i^T\boldsymbol{x}_j)^2 - 2\{n(n-1)(n-2)\}^{-1}\sum_{i\neq j\neq k}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_j\boldsymbol{x}_j^T\boldsymbol{x}_k + \{n(n-1)(n-2)(n-3)\}^{-1}\sum_{i\neq j\neq k\neq l}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_j\boldsymbol{x}_k^T\boldsymbol{x}_l$. Note that the asymptotic variance of $W_{n(Z)}$ is equivalent to that of W_n under (A-i). However, the computational cost of $W_{n(Z)}$ is written by the order, $O(n^4p)$. Contrary to that, the computational cost of W_n is $O(n^2p)$ by substituting previously calculated $\overline{\boldsymbol{x}}_{n(i)(k)}$s in (9). In conclusion, the ECDM methodology is an efficient method to construct an unbiased estimator in non-Gaussian situations.

3. Sample size determination to control both size and power

We are interested in designing a test of (1) having size α and power no less than $1-\beta$ when $||\boldsymbol{\sigma}||^2/\{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\} \geq \Delta_L$, where $\alpha \in (0,1/2)$, $\beta \in (0,1/2)$ and Δ_L (>0) are prespecified constants. We note that $||\boldsymbol{\sigma}||^2/\{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\} \in [0,1]$. We emphasize that $||\boldsymbol{\sigma}||^2/\{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\}$ represents a contribution of $x_{j(*)}$ to $\boldsymbol{x}_j$. See Remark 8. We assume that $\Delta_L = o\{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}/\mathrm{tr}(\boldsymbol{\Sigma})\}$ and $\liminf_{p\to\infty} p\Delta_L > 0$.

3.1. Sample size determination

We consider n satisfying

$$\sqrt{\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})} \leq \frac{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\Delta_L}{z_\alpha + z_\beta} \quad \text{when} \quad \frac{||\boldsymbol{\sigma}||^2}{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})} \leq \Delta_L$$

under (A-i) to (A-iii). Then, one would find the sample size such as

$$n \geq \frac{(z_\alpha + z_\beta)}{\Delta_L\mathrm{tr}(\boldsymbol{\Sigma})}\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)} + \frac{2(z_\alpha+z_\beta)^2}{\sigma_*^4\Delta_L^2\mathrm{tr}(\boldsymbol{\Sigma})^2}\Big(\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma} + ||\boldsymbol{\sigma}||^4 + \sum_{i=1}^{r}(M_i-2)c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2\Big) \quad (= C,\ \text{say}). \tag{12}$$

We consider testing the hypothesis (1) by

$$\text{rejecting } H_0 \Longleftrightarrow \widehat{T}_{n,\boldsymbol{\sigma}} > \frac{S_{n(*)}\mathrm{tr}(\boldsymbol{S}_n)\Delta_L z_\alpha}{z_\alpha + z_\beta}. \tag{13}$$

Note that $C \to \infty$, namely, $n \to \infty$ as $p \to \infty$ from the fact that $\Delta_L = o\{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}/\mathrm{tr}(\boldsymbol{\Sigma})\}$ as $p \to \infty$. Then, we have the following theorem.

Theorem 3.1. *Under (A-i) to (A-iii), the test by (12)-(13) has that*

$$\limsup_{p\to\infty} \textit{size} \le \alpha \quad \textit{and} \quad \liminf_{p\to\infty} \textit{power} \ge 1-\beta \quad \textit{when} \quad \frac{||\boldsymbol{\sigma}||^2}{\sigma_*^2 tr(\boldsymbol{\Sigma})} \ge \Delta_L.$$

Remark 7. Note that (A-ii) implies $\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}/\mathrm{tr}(\boldsymbol{\Sigma}) \to 0$ as $p \to \infty$ from (A.5) in Appendix. Then, it holds as $p \to \infty$ that $C/p \to 0$ under $\limsup_{p\to\infty} ||\boldsymbol{\sigma}||^2/\{\sigma_*^2\Delta_L \mathrm{tr}(\boldsymbol{\Sigma})\} < \infty$ and $\liminf_{p\to\infty} p\Delta_L > 0$.

Remark 8. We consider a multivariate linear regression model such as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\Upsilon} + \boldsymbol{E}.$$

Here, $\boldsymbol{Y} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_n]^T$ is an $n \times p$ response matrix, $\boldsymbol{X} = [\mathbf{1}, (x_{1(*)}, ..., x_{n(*)})^T]$ is an $n \times 2$ fixed design matrix having $\mathbf{1} = (1, ..., 1)^T$, and $\boldsymbol{\Upsilon}$ is a $2 \times p$ parameter matrix. The n rows of $\boldsymbol{E}$ are independent and identically distributed as a p-variate distribution with mean vector zero and covariance matrix $\boldsymbol{\Sigma}$. We assume that the fourth moments of each variable in $\boldsymbol{E}$ are uniformly bounded. A squared multiple correlation coefficient is given by $R^2 = ||\boldsymbol{s}_{n(*)}||^2/\{S_{n(*)}\mathrm{tr}(\boldsymbol{S}_n)\}$, where $\boldsymbol{s}_{n(*)}$, $S_{n(*)}$ and $\boldsymbol{S}_n$ are defined in (4). We assume that $S_{n(*)} \to \sigma_*^2$ and $\boldsymbol{s}_{n(*)} \to \boldsymbol{\sigma}$ in probability as $n \to \infty$. Note that $\mathrm{tr}(\boldsymbol{S}_n) \to \mathrm{tr}(\boldsymbol{\Sigma})$ in probability as $n \to \infty$. Then, it holds as $n \to \infty$ that $R^2 \to ||\boldsymbol{\sigma}||^2/\{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\}$ in probability. Thus one can apply the correlation test procedure to a test whether $R^2 = 0$ or $R^2 \neq 0$.

3.2. Two-stage procedure

Since C includes unknown parameters, it is necessary to estimate C in (12) with some pilot samples. However, it is very difficult to estimate $\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma}$ and $\sum_{i=1}^r (M_i - 1)c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2$. Hence, from the fact that

$$\frac{\sum_{i=1}^r c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2}{\sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma})} \le \frac{\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma}}{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})} \le \lambda_1\Delta_L \le \mathrm{tr}(\boldsymbol{\Sigma}^4)^{1/4}\Delta_L = o\{\mathrm{tr}(\boldsymbol{\Sigma}^2)^{1/2}\Delta_L\}$$

when $||\boldsymbol{\sigma}||^2/\{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\} \le \Delta_L$ under (A-ii), we modify C as follows:

$$C = \frac{(z_\alpha + z_\beta)}{\Delta_L\mathrm{tr}(\boldsymbol{\Sigma})}\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)\{1+o(1)\} + 2(z_\alpha+z_\beta)^2\frac{||\boldsymbol{\sigma}||^4}{\sigma_*^4\Delta_L^2\mathrm{tr}(\boldsymbol{\Sigma})^2}} \tag{14}$$

$$\approx \frac{(z_\alpha + z_\beta)}{\Delta_L\mathrm{tr}(\boldsymbol{\Sigma})}\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2) + 2\eta(z_\alpha+z_\beta)^2} \quad (= C_\star,\ \text{say}), \tag{15}$$

where $\eta \in [0,1]$ is a chosen constant. See Remark 9 for a choice of η. Note that $C_\star/C \to 1$ as $p \to \infty$ when $||\boldsymbol{\sigma}||^2/\{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\} \le \Delta_L$ under (A-ii). We propose a two-stage test procedure in order to estimate $C_\star$ assuring the prespecified accuracy. We proceed with the following two steps:

1. Choose m (≥ 4) such as

$$\frac{m}{C_\star} \le 1, \quad \frac{C_\star}{m^2} \to 0 \ \text{ and } \ \frac{C_\star}{m}\frac{\mathrm{tr}(\boldsymbol{\Sigma}^4)}{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2} \to 0 \ \text{ as } \ p \to \infty. \tag{16}$$

Note that m satisfying $m/C_\star \to c \in (0,1)$ as $p \to \infty$ holds (16) under (A-ii). Also, note that $\mathrm{Var}_{\boldsymbol{\theta}}\{\mathrm{tr}(\boldsymbol{S}_m)/\mathrm{tr}(\boldsymbol{\Sigma})\} = o(C_\star^{-1})$ and $\mathrm{Var}_{\boldsymbol{\theta}}\{W_m/\mathrm{tr}(\boldsymbol{\Sigma}^2)\} = o(C_\star^{-1})$ under (A-i)-(A-ii) and (16). Take pilot samples, $\boldsymbol{x}_{j(*)} = (\boldsymbol{x}_j^T, x_{j(*)})^T$, $j = 1, ..., m$. Then, calculate $\boldsymbol{S}_m$ and W_m according to (4) and (9). Define the total sample size by

$$N = \max\Big\{m, \ \Big\lceil \frac{(z_\alpha + z_\beta)}{\Delta_L \mathrm{tr}(\boldsymbol{S}_m)}\sqrt{2W_m} + 2\eta(z_\alpha + z_\beta)^2 \ \Big\rceil\Big\}. \tag{17}$$

2. If $N = m$, do not take any additional samples. If $N > m$, take additional samples, $\boldsymbol{x}_{j(*)}$, $j = m+1, ..., N$. By combining the initial samples and the additional samples, calculate $S_{N(*)}$, $\boldsymbol{S}_N$ and $\widehat{T}_{N,\boldsymbol{\sigma}}$ according to (4) and (6). Under (A-i)-(A-iii), from the fact that $C = C_\star - 2\eta(z_\alpha + z_\beta)^2$ when $\boldsymbol{\rho} = \boldsymbol{0}$, it holds that

$$\sqrt{\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{C_{(\star)},\boldsymbol{\sigma}})} = \frac{\{1 - 2\eta(z_\alpha + z_\beta)^2/C_{(\star)}\}\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\Delta_L}{z_\alpha + z_\beta}\{1 + o(1)\}$$

when $\boldsymbol{\rho} = \boldsymbol{0}$, where $C_{(\star)} = \lceil C_\star \rceil$. Then, test the hypothesis (1) by

$$\text{rejecting } H_0 \Longleftrightarrow \widehat{T}_{N,\boldsymbol{\sigma}} > \frac{\{1 - 2\eta(z_\alpha + z_\beta)^2/N\}S_{N(*)}\mathrm{tr}(\boldsymbol{S}_N)\Delta_L z_\alpha}{z_\alpha + z_\beta}. \tag{18}$$

We have the following theorem.

Theorem 3.2. *Under (A-i) to (A-iii), the test by (18) with (16)-(17) has that*

$$\lim_{p\to\infty} \textit{size} = \alpha \quad \textit{and} \quad \liminf_{p\to\infty} \textit{power} \ge 1 - \beta \quad \textit{when} \quad \frac{||\boldsymbol{\sigma}||^2}{\sigma_*^2 \textit{tr}(\boldsymbol{\Sigma})} \ge \Delta_L.$$

Remark 9. When the lower bound is attained, namely $||\boldsymbol{\sigma}||^2 = \sigma_*^2 \mathrm{tr}(\boldsymbol{\Sigma})\Delta_L$, we claim from (14) that $\{\sigma_*^2 \mathrm{tr}(\boldsymbol{\Sigma})\Delta_L\ /(z_\alpha + z_\beta)\}/\{\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{C_{(\star)},\boldsymbol{\sigma}})\}^{1/2} = \{1 + o(1)\}C_{(\star)}/C = \{1 + o(1)\}/\{1 + 2(1-\eta)(z_\alpha + z_\beta)^2/C_{(\star)}\} \to 1$ under (A-i) to (A-iii). Let $\zeta = 2(z_\alpha + z_\beta)^2/C_{(\star)}$. Then, from Theorem 2.1 and (18), it holds that

$$\frac{\{1-\eta\zeta\}\sigma_*^2 \mathrm{tr}(\boldsymbol{\Sigma})\Delta_L z_\alpha}{(z_\alpha + z_\beta)\{\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{C_{(\star)},\boldsymbol{\sigma}})\}^{1/2}} - \frac{||\boldsymbol{\sigma}||^2}{\{\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{C_{(\star)},\boldsymbol{\sigma}})\}^{1/2}}$$
$$= \frac{\{1-\eta\zeta\}z_\alpha - z_\alpha - z_\beta}{1+(1-\eta)\zeta}\{1+o(1)\} = \frac{-z_\beta(1+\eta\zeta z_\alpha/z_\beta)}{1+(1-\eta)\zeta}\{1+o(1)\} \to -z_\beta$$

when $||\boldsymbol{\sigma}||^2 = \sigma_*^2 \mathrm{tr}(\boldsymbol{\Sigma})\Delta_L$. Thus one may choose η such that $1 + \eta\zeta z_\alpha/z_\beta = 1 + (1-\eta)\zeta$, that is $\eta = z_\beta/(z_\alpha + z_\beta)$.

Remark 10. It holds as $p \to \infty$ that $N/C_\star = 1 + o_p(1)$ and $C_\star/p \to 0$; that is in the HDLSS situation in the sense that $N/p = o_p(1)$.

Remark 11. One can claim that $\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}/\mathrm{tr}(\boldsymbol{\Sigma}) \geq p^{-1/2}$, where the equality holds only when $\lambda_1 = \cdots = \lambda_p$. For the cases in Remark 1, it holds that $\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}/\mathrm{tr}(\boldsymbol{\Sigma}) = O(p^{-1/2})$. Thus for those cases, one may choose a pilot sample size by

$$m = \max\left\{4,\ \left\lceil \frac{(z_\alpha + z_\beta)\sqrt{2}}{\Delta_L\sqrt{p}} + 2\eta(z_\alpha + z_\beta)^2 \right\rceil\right\}.$$

Then, (16) holds under (A-ii).

Remark 12. One may choose m (≥ 4) such as $m/C_\star > 1$. Then, the assertion in Theorem 3.2 is still claimed. However, it may cause over-sampling in the sense that $N/C_\star > 1$ w.p.1.

3.3. Moderate sample performances

In order to study the performance of the two-stage test procedure given by (18) with (16)-(17), we used computer simulations. We fixed $\Delta_L = 5/p$. Our goal was to construct a test having size $\alpha = 0.05$ and power no less than $1-\beta = 0.9$ when $||\boldsymbol{\sigma}||^2/\{\sigma_*^2 \mathrm{tr}(\boldsymbol{\Sigma})\} \geq \Delta_L$. We considered a non-Gaussian case by setting $r = p$, $\boldsymbol{\Gamma} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}$ and $w_{ij} = (8/10)^{1/2} v_{ij}$ in (2), where v_{ij}, $i = 1, ..., p$ $(j = 1, 2, ...)$ are independently distributed as t-distribution with 10 degrees of freedom. Note that $E(w_{ij}) = 0$, $E(w_{ij}^2) = 1$, and (A-i) holds.

Independent of v_{ij}s, we generated v_{j*}s independently from the pseudorandom t-distribution with 10 degrees of freedom. We set $w_{j*} = (8/10)^{1/2} v_{j*}$ $(j = 1, 2, ...)$ so as to satisfy (A-iii). We considered $\sigma_*^2 = 1$ and $\boldsymbol{\Sigma} = \boldsymbol{B}(\rho^{|i-j|^{1/3}})\boldsymbol{B}$ having $\rho \in (0, 1)$ and

$$\boldsymbol{B} = \mathrm{diag}(\sqrt{0.5 + 1/(p+1)}, \sqrt{0.5 + 2/(p+1)}, ..., \sqrt{0.5 + p/(p+1)}).$$

Note that $\mathrm{tr}(\boldsymbol{\Sigma}) = p$. We set $m = \lceil C_\star/2 \rceil$. We considered two choices of η as $\eta = z_\beta/(z_\alpha + z_\beta)$ from Remark 9 and $\eta = 1$. We considered the following four cases when $p = 500$ and 1000: (a) $\rho = 0.3$ and $\eta = z_\beta/(z_\alpha + z_\beta)$; (b) $\rho = 0.3$ and $\eta = 1$; (c) $\rho = 0.5$ and $\eta = z_\beta/(z_\alpha + z_\beta)$; and (d) $\rho = 0.5$ and $\eta = 1$.

In Table 1, we summarized the findings obtained by averaging the outcomes from 4000 ($= R$, say) replications, where the first 2000 replications were generated for $\boldsymbol{\rho} = \mathbf{0}$ by setting as $c_1 = \cdots = c_p = 0$ and $c_* = 1$ in (3), and the last 2000 replications were generated for $\boldsymbol{\rho} \neq \mathbf{0}$ by setting as $c_g = \sqrt{5/\lambda_g}$, $c_* = \sqrt{1 - c_g^2}$ and the other c_is are 0 (i.e., $||\boldsymbol{\sigma}||^2 = c_g^2 \lambda_g = 5$ and $||\boldsymbol{\sigma}||^2/\{\sigma_*^2 \mathrm{tr}(\boldsymbol{\Sigma})\} = 5/p$) in (3). Here, we set $g = 5$ for $\rho = 0.3$ and $g = 10$ for $\rho = 0.5$. Under a fixed scenario, suppose that the rth replication ends with $N = n_r$ observations given by (17) and the test result given by (18). We defined $P_r = 1$ (or 0) accordingly as $H_0 :\ \boldsymbol{\rho} = \mathbf{0}$ was falsely rejected (or not) and $H_1 :\ \boldsymbol{\rho} \neq \mathbf{0}$ was falsely rejected (or not). We defined $\overline{\alpha} = (R/2)^{-1} \sum_{r=1}^{R/2} P_r$ to estimate the size and $1 - \overline{\beta} = 1 - (R/2)^{-1} \sum_{r=R/2+1}^{R} P_r$ to estimate the power when $||\boldsymbol{\sigma}||^2/\{\sigma_*^2 \mathrm{tr}(\boldsymbol{\Sigma})\} = \Delta_L$, while their estimated standard errors, $s(\overline{\alpha})$ and $s(\overline{\beta})$, were given by $s^2(\overline{\alpha}) = (R/2)^{-1}\overline{\alpha}(1 - \overline{\alpha})$ and $s^2(\overline{\beta}) = (R/2)^{-1}\overline{\beta}(1 - \overline{\beta})$. We also defined $\overline{n} = R^{-1} \sum_{r=1}^{R} n_r$ and $\mathrm{Var}(n) = (R-1)^{-1} \sum_{r=1}^{R} (n_r - \overline{n})^2$.

When $\rho = 0.3$, we observed that the test by (18) with (16)-(17) provides good performances. Especially, the test having $\eta = z_\beta/(z_\alpha + z_\beta)$ gave adequate performances about the target rates, $\alpha = 0.05$ and $\beta = 0.1$. On the other hand, the test having $\eta = 1$ satisfied the target rates excessively by taking samples more than needs. When $\rho = 0.5$, we observed that the test having $\eta = z_\beta/(z_\alpha + z_\beta)$ gave error rates a little upper than the target rates. Note that, for $p = 1000$, $\mathrm{tr}(\boldsymbol{\Sigma}^4)/\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 = 0.094$ when $\rho = 0.5$, while $\mathrm{tr}(\boldsymbol{\Sigma}^4)/\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 = 0.011$ when $\rho = 0.3$. The slightly low accuracy may be attributed to a slow convergence in (A-ii) when $\rho = 0.5$. On the other hand, the test having $\eta = 1$ gave good performances even when $\rho = 0.5$.

Table 1: Required sample size, and the size and power by (18) with (16)-(17).

$C_\star$	$\overline{n}$	$\overline{n}-C_\star$	$\mathrm{Var}(n)$	$\overline{\alpha}$	$s(\overline{\alpha})$	$1-\overline{\beta}$	$s(\overline{\beta})$
When $p=500$							
(a) $\rho=0.3$ and $\eta=z_\beta/(z_\alpha+z_\beta)$: m=16							
31.72	32.11	0.39	4.59	0.056	0.00514	0.903	0.00663
(b) $\rho=0.3$ and $\eta=1$: m=21							
41.35	41.76	0.42	2.69	0.052	0.00494	0.982	0.00301
(c) $\rho=0.5$ and $\eta=z_\beta/(z_\alpha+z_\beta)$: m=25							
49.74	49.64	-0.1	23.64	0.064	0.00547	0.898	0.00678
(d) $\rho=0.5$ and $\eta=1$: m=30							
59.37	59.45	0.08	19.46	0.061	0.00535	0.955	0.00466
When $p=1000$							
(a) $\rho=0.3$ and $\eta=z_\beta/(z_\alpha+z_\beta)$: m=21							
41.8	42.37	0.57	4.67	0.059	0.00527	0.894	0.00688
(b) $\rho=0.3$ and $\eta=1$: m=26							
51.43	52.03	0.6	3.16	0.057	0.00518	0.957	0.00454
(c) $\rho=0.5$ and $\eta=z_\beta/(z_\alpha+z_\beta)$: m=34							
67.93	68.02	0.09	23.0	0.06	0.00531	0.869	0.00754
(d) $\rho=0.5$ and $\eta=1$: m=39							
77.56	77.57	0.01	18.54	0.059	0.00527	0.916	0.0062

4. Multiple testing procedures

In this section, we propose multiple testing procedures for high-dimensional data. Suppose we have i.i.d. $p+K$-variate data vectors, $\boldsymbol{x}_{j(*)}=(\boldsymbol{x}_j^T, x_{1j(*)},...,x_{Kj(*)})^T$, $j=1,...,n$, where $\boldsymbol{x}_j$ is defined in Section 1 and K is an integer ≥ 2. Here, $x_{ij(*)}$ has unknown mean, μ_{i*}, and unknown variance, $\sigma_{i*}^2\in(0,\infty)$, for each i $(=1,...,K)$. Let $\boldsymbol{\theta}_K=(\mu_{1*},...,\mu_{K*},\sigma_{1*}^2,...,\sigma_{K*}^2,\boldsymbol{\mu},\boldsymbol{\Sigma})$. We denote the covariance vector between $\boldsymbol{x}_j$ and $x_{ij(*)}$ by $\mathrm{Cov}_{\boldsymbol{\theta}_K}(\boldsymbol{x}_j,x_{ij(*)})=\boldsymbol{\sigma}_i$ $(i=1,...,K)$. We denote the correlation coefficient vector between $\boldsymbol{x}_j$ and $x_{ij(*)}$ by $\mathrm{Corr}_{\boldsymbol{\theta}_K}(\boldsymbol{x}_j,x_{ij(*)})=\boldsymbol{\rho}_i$ $(i=1,...,K)$.

Let

$$x_{ij(*)}=c_{i*}w_{ij*}+\sum_{i'=1}^{r}c_{ii'}w_{i'j}+\mu_{i*},\ i=1,...,K, \tag{19}$$

where $w_{i'j}$s are defined in (3), and w_{ij*} $(i=1,...,K)$ is a random variable such that $E(w_{ij*})=0$, $E(w_{ij*}^2)=1$ and $E(w_{i'j}w_{ij*})=0$ for $i'=1,...,r$. Here, c_{i*} and $c_{ii'}$s are constants such that $c_{i*}^2+\sum_{i'=1}^{r}c_{ii'}^2=\sigma_{i*}^2$. Note that $\sum_{i'=1}^{r}c_{ii'}\boldsymbol{\gamma}_{i'}=\boldsymbol{\sigma}_i$. We assume the following assumption for w_{ij*} as necessary:

(A-v) The fourth moment of w_{ij*} is bounded, and w_{ij*} and $\boldsymbol{w}_j$ are independent for $i=1,...,K$.

We consider a multiple test of the correlation between $\boldsymbol{x}_j$ and $x_{ij(*)}$s by

$$H_{0i}:\boldsymbol{\rho}_i=\boldsymbol{0}\quad \text{vs.}\quad H_{1i}:\boldsymbol{\rho}_i\neq\boldsymbol{0}\quad \text{for } i=1,...,K. \tag{20}$$

Our interest is to select a set of significant correlated variables such as $\boldsymbol{D}=\{i|\ i\in\{1,...,K\}\text{ such that }\boldsymbol{\rho}_i\neq\boldsymbol{0}\}$. We apply the proposed correlation testing procedure to the multiple test. A test procedure $\widehat{\boldsymbol{D}}$ maps the data into subsets of $\{1,...,K\}$.

4.1. Multiple test of correlations to control family-wise error rate

We are interested in designing $\widehat{\boldsymbol{D}}$ such that the *family-wise error rate* (FWER) is $\leq\alpha$, i.e.

$$P_{\boldsymbol{\theta}_K}(\boldsymbol{D}^c\cap\widehat{\boldsymbol{D}}\neq\emptyset)\leq\alpha.$$

Let

$$\widehat{T}_{n,\boldsymbol{\sigma}(i)} = \frac{2u_n}{n(n-1)} \sum_{j<l}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(1)(j+l)})^T (\boldsymbol{x}_l - \overline{\boldsymbol{x}}_{n(2)(j+l)})$$
$$\times (x_{ij(*)} - \overline{x}_{in(1*)(j+l)})(x_{il(*)} - \overline{x}_{in(2*)(j+l)}) \tag{21}$$
$$\text{and} \quad S_{in(*)} = \frac{\sum_{j=1}^{n} (x_{ij(*)} - \overline{x}_{in(*)})^2}{n-1}, \ i = 1, ..., K,$$

where $\overline{x}_{in(1*)(k)} = \sum_{j \in \boldsymbol{V}_{n(1)(k)}} x_{ij(*)}/n_{(1)}$, $\overline{x}_{in(2*)(k)} = \sum_{j \in \boldsymbol{V}_{n(2)(k)}} x_{ij(*)}/n_{(2)}$, $k = 3, ..., 2n-1$, and $\overline{x}_{in(*)} = \sum_{j=1}^{n} x_{ij(*)}/n$. Then, from Corollary 2.2, under (A-i), (A-ii) and (A-v), it holds as $p \to \infty$ and $n \to \infty$ that

$$\frac{\widehat{T}_{n,\boldsymbol{\sigma}(i)}}{S_{in(*)}\sqrt{2W_n/n}} \Rightarrow N(0,1) \quad \text{for } i \in \boldsymbol{D}^c.$$

Here, from Bonferroni's method, we test the hypotheses (20) by

$$\text{rejecting } H_{0i} \Longleftrightarrow \frac{\widehat{T}_{n,\boldsymbol{\sigma}(i)}}{S_{in(*)}\sqrt{2W_n/n}} > z_{\alpha/K}. \tag{22}$$

Let $\widehat{\boldsymbol{D}} = \{i | \ i \in \{1, ..., K\} \text{ rejecting } H_{0i}\}$. Then, we have as $p \to \infty$ and $n \to \infty$ that

$$\limsup P_{\boldsymbol{\theta}_K}(\boldsymbol{D}^c \cap \widehat{\boldsymbol{D}} \neq \emptyset) \leq \alpha$$

under (A-i), (A-ii) and (A-v).

Remark 13. By using the asymptotic p-value given by

$$P_i = 1 - \Phi\left(\frac{\widehat{T}_{n,\boldsymbol{\sigma}(i)}}{S_{in(*)}\sqrt{2W_n/n}}\right), \quad i = 1, ..., K,$$

one may apply the Bonferroni-Holm method given by Holm [18] or the false discovery rate (FDR) controlling procedure given by Benjamini and Hochberg [7] and Benjamini and Yekutieli [8].

4.2. Multiple test of correlations to control both FWER and power

We consider a test of (20) having FWER $\leq \alpha$ and power

$$P_{\boldsymbol{\theta}_K}(\boldsymbol{D} \subseteq \widehat{\boldsymbol{D}}) \geq 1 - \beta \quad \text{when } \min_{i \in \boldsymbol{D}} \frac{||\boldsymbol{\sigma}_i||^2}{\sigma_{i*}^2 \text{tr}(\boldsymbol{\Sigma})} \geq \Delta_L,$$

where $\alpha \in (0, 1/2)$, $\beta \in (0, 1/2)$ and Δ_L (> 0) are prespecified constants. We assume $\Delta_L = o\{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}/\mathrm{tr}(\boldsymbol{\Sigma})\}$ and $\liminf_{p\to\infty} p\Delta_L > 0$. Then, we propose a two-stage test procedure based on the following two steps:

1. Choose $m(\geq 4)$ satisfying (16). Take pilot samples $\boldsymbol{x}_{j(*)} = (\boldsymbol{x}_j^T, x_{1j(*)}, ..., x_{Kj(*)})^T$, $j = 1, ..., m$. Then, calculate $\boldsymbol{S}_m$ and W_m according to (4) and (9). Define the total sample size by

$$N = \max\Big\{m,\ \Big\lceil \frac{(z_{\alpha/K} + z_{\beta/K})}{\Delta_L \mathrm{tr}(\boldsymbol{S}_m)}\sqrt{2W_m} + 2\eta(z_{\alpha/K} + z_{\beta/K})^2 \Big\rceil\Big\}, \tag{23}$$

where $\eta \in [0, 1]$ is a chosen constant. See Remark 14 for a choice of η.

2. If $N = m$, do not take any additional samples. If $N > m$, take additional samples, $\boldsymbol{x}_{j(*)}$, $j = m + 1, ..., N$. By combining the initial samples and the additional samples, calculate $\boldsymbol{S}_N$, $S_{iN(*)}$ and $\widehat{T}_{N,\boldsymbol{\sigma}(i)}$, $i = 1, ..., K$, according to (4) and (21). Then, test the hypotheses (20) by

$$\text{rejecting } H_{0i} \Longleftrightarrow \widehat{T}_{N,\boldsymbol{\sigma}(i)} > \frac{\{1 - 2\eta(z_{\alpha/K} + z_{\beta/K})^2/N\} S_{iN(*)}\mathrm{tr}(\boldsymbol{S}_N)\Delta_L z_{\alpha/K}}{z_{\alpha/K} + z_{\beta/K}}. \tag{24}$$

Then, we have the following theorem.

Theorem 4.1. *Under (A-i), (A-ii) and (A-v), the test by (24) with (23) has that*

$$(i)\quad \limsup_{p\to\infty} P_{\boldsymbol{\theta}_K}(\boldsymbol{D}^c \cap \widehat{\boldsymbol{D}} \neq \emptyset) \leq \alpha;$$

$$(ii)\quad \liminf_{p\to\infty} P_{\boldsymbol{\theta}_K}(\boldsymbol{D} \subseteq \widehat{\boldsymbol{D}}) \geq 1 - \beta \quad \textit{when } \min_{i\in\boldsymbol{D}} \frac{||\boldsymbol{\sigma}_i||^2}{\sigma_{i*}^2 tr(\boldsymbol{\Sigma})} \geq \Delta_L.$$

Remark 14. Note that $N/p = o_p(1)$ under (A-i) and (A-ii) from the facts that $\liminf_{p\to\infty} p\Delta_L > 0$ and (A.5) in Appendix. From Remark 9, one may define η as $\eta = z_{\beta/K}/(z_{\alpha/K} + z_{\beta/K})$.

5. Data analysis

In this section, we demonstrate how the test procedures perform in actual data analyses by using two microarray data sets.

5.1. T-cell acute lymphoblastic leukemia

We analyzed gene expression data of T-cell acute lymphoblastic leukemia (T-ALL) given by Chiaretti et al. [10] in which the data set consists of 12625 genes and 33 ($= n$) samples. Note that the expression measures were obtained by using the three-step robust multichip average (RMA) preprocessing method. Refer to Pollard et al. [23] as well for the details.

Chiaretti et al. [10] identified 3 predictive genes, *TTK*, *AHNAK* and *CD2*, to distinguish the patients according to disease outcomes. On the other hand, Gottardo et al. [15] identified 3 predictive genes, *NOTCH2*, *BTG3* and *CFLAR*. We denoted these 6 predictive genes, $(TTK, AHNAK, CD2, NOTCH2, BTG3, CFLAR)$, by $x_{ij(*)},\ i = 1, ..., K(= 6)$. We denoted the remaining 12619 ($= p$) genes by $\boldsymbol{x}_j$. We considered a multiple testing to see whether the predictive genes have a significant influence of the other genes' expression or not. Let $\alpha = 0.05$. Our goal was to find variables i's having $\boldsymbol{\rho}_i \neq \mathbf{0}$ with respect to FWER given by

$$P_{\boldsymbol{\theta}_K}(\boldsymbol{D}^c \cap \widehat{\boldsymbol{D}} \neq \boldsymbol{\emptyset}) \leq 0.05.$$

We applied the multiple test given by (22). According to (21), we calculated $\widehat{T}_{n,\boldsymbol{\sigma}(1)} = 170.92$ (TTK), $\widehat{T}_{n,\boldsymbol{\sigma}(2)} = 60.33$ $(AHNAK)$, $\widehat{T}_{n,\boldsymbol{\sigma}(3)} = 44.74$ $(CD2)$, $\widehat{T}_{n,\boldsymbol{\sigma}(4)} = 1.03$ $(NOTCH2)$, $\widehat{T}_{n,\boldsymbol{\sigma}(5)} = 14.24$ $(BTG3)$ and $\widehat{T}_{n,\boldsymbol{\sigma}(6)} = 5.24$ $(CFLAR)$ by using the data set with $n = 33$. With the help of the multiple test given by (22) with $z_{\alpha/K} = 2.394$, we selected a set of significant genes by

$$\widehat{\boldsymbol{D}} = \{1, 2, 3, 6\},$$

guaranteeing the FWER. The selected 4 genes were $(TTK, AHNAK, CD2, CFLAR)$. We observed that three predictive genes given by Chiaretti et al. [10] and one predictive gene given by Gottardo et al. [15] have a significant influence of the other genes' expression. On the other hand, the remaining two predictive genes given by Gottardo et al. were considered to be unrelated to the other genes' expression. Those 2 predictive genes, $(NOTCH2, BTG3)$, may distinguish the patients according to disease outcomes without a influence of the other genes' expression.

5.2. Arabidopsis thaliana

We analyzed gene expression data of Arabidopsis thaliana given by Wille et al. [27] in which the data set consists of 118 samples having 39 ($= K$)

isoprenoid genes and 795 $(=p)$ additional genes. All data were logarithmic transformed and denoted by $x_{ij(*)},\ i=1,...,K$, for the isoprenoid genes and by $\boldsymbol{x}_j$ for the additional genes. Wille et al. [27] considered a genetic network between the two gene sets. We considered a multiple testing to select a significant set of associated genes from among isoprenoid genes. Specifically, we were interested in finding the interplay between $\boldsymbol{x}_j$ and each $x_{ij(*)}$. Let $\alpha=0.05$, $\beta=0.1$ and $\Delta_L=0.1$. Our goal was to find variables i's having $\boldsymbol{\rho}_i\neq\mathbf{0}$ with FWER given by $P_{\boldsymbol{\theta}_K}(\boldsymbol{D}^c\cap\widehat{\boldsymbol{D}}\neq\boldsymbol{\emptyset})\leq 0.05$ and power given by $P_{\boldsymbol{\theta}_K}(\boldsymbol{D}\subseteq\widehat{\boldsymbol{D}})\geq 0.9$ when $\min_{i\in\boldsymbol{D}}||\boldsymbol{\sigma}_i||^2/\{\sigma_{i*}^2\mathrm{tr}(\boldsymbol{\Sigma})\}\geq 0.1$. We applied the two-stage test procedure given by (24) with (23) to the inference. We set $\eta=z_{\beta/39}/(z_{\alpha/39}+z_{\beta/39})$ from Remark 14. From Remark 11, we calculated the pilot sample size as

$$m=\max\left\{4,\ \left\lceil\frac{(z_{\alpha/39}+z_{\beta/39})\sqrt{2}}{0.1\sqrt{p}}+2z_{\beta/39}(z_{\alpha/39}+z_{\beta/39})\right\rceil\right\}=36.$$

So, we took the first 36 samples as a pilot sample. We calculated $\mathrm{tr}(\boldsymbol{S}_m)=440$ and $W_m=13935$ according to (4) and (9), respectively. Then, from (23), we had the total sample size as

$$N=\max\left\{m,\ \left\lceil\frac{(z_{\alpha/39}+z_{\beta/39})\sqrt{2}}{0.1\times 440}\sqrt{13935}+2z_{\beta/39}(z_{\alpha/39}+z_{\beta/39})\right\rceil\right\}=55.$$

Thus we took the next 19 $(=55-36)$ samples. Then, we calculated $\widehat{T}_{N,\boldsymbol{\sigma}(i)}$, $\boldsymbol{S}_N$ and $S_{iN(*)}$, $i=1,...,39$, according to (4) and (21). By using the multiple test given by (24), we selected a set of significant genes by

$$\widehat{\boldsymbol{D}}=\{1,...,39\}\setminus\{6,7,13,14,15,16,17,20\},$$

guaranteeing both the FWER and the power. Thus we selected 31 isoprenoid genes. We considered a high-dimensional linear regression model:

$$\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{\Upsilon}+\boldsymbol{E},$$

where $\boldsymbol{Y}$ is an $n\times p$ response matrix, $\boldsymbol{X}$ is an $n\times K'$ fixed design matrix, and $\boldsymbol{\Upsilon}$ is a $K'\times p$ parameter matrix. The n rows of $\boldsymbol{E}$ are independent and identically distributed as a p-variate distribution with mean vector zero. When $K'=2$, see Remark 8. Let $x_{j(1*)},...,x_{j(31*)}$, be the jth sample of the 31 selected isoprenoid genes in $\widehat{\boldsymbol{D}}$. Let $\boldsymbol{x}_{(j)}=(1,x_{j(1*)},...,x_{j(31*)})^T$,

$j=1,...,n$. We set $\boldsymbol{Y}=[\boldsymbol{x}_1,...,\boldsymbol{x}_n]^T$ and $\boldsymbol{X}=[\boldsymbol{x}_{(1)},...,\boldsymbol{x}_{(n)}]^T$ with $K'=32$. We noted that the standard elements of $\boldsymbol{\Upsilon}$ are path coefficients from the isoprenoid genes to the additional genes. By using the observed samples of size $n=55$ as a training data set, we obtained the least squared estimator of $\boldsymbol{\Upsilon}$ by $\widehat{\boldsymbol{\Upsilon}}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$. We investigated prediction accuracy of the regression with $\widehat{\boldsymbol{\Upsilon}}$ by using the remaining samples of size 63 $(=118-55)$ as a test data set. We denoted the test samples by $x_{j(i*)}$ $(i=1,...,31)$ and $\boldsymbol{x}_j$, $j=56,...,118$. We considered the prediction mean squared error (PMSE) by $E(||\boldsymbol{x}_j-\widehat{\boldsymbol{\Upsilon}}^T\boldsymbol{x}_{(j)}||^2|\widehat{\boldsymbol{\Upsilon}})$. By using the test samples $x_{j(i*)}$ $(i=1,...,31)$ and $\boldsymbol{x}_j$, $j=56,...,118$, we applied the bias-corrected and accelerated (BCa) bootstrap by Efron [12]. Then, we constructed 95% confidence interval (CI) of the PMSE by $[704.2, 955.5]$ from 10000 replications. We also calculated the sample mean of the PMSE by 809.5.

On the other hand, we considered the PMSE for the full isoprenoid genes by $E(||\boldsymbol{x}_j-\widehat{\boldsymbol{\Upsilon}}_f^T\boldsymbol{x}_{f(j)}||^2|\widehat{\boldsymbol{\Upsilon}}_f)$, where $\widehat{\boldsymbol{\Upsilon}}_f=(\boldsymbol{X}_f^T\boldsymbol{X}_f)^{-1}\boldsymbol{X}_f^T\boldsymbol{Y}$ with $\boldsymbol{X}_f=[\boldsymbol{x}_{f(1)},...,\boldsymbol{x}_{f(55)}]^T$ and $\boldsymbol{x}_{f(j)}=(1,x_{1j(*)},...,x_{39j(*)})^T$, $j=1,...,55$. Then, similarly to above, we constructed 95% CI of the PMSE by $[897.9, 1217.8]$. We also calculated the sample mean of the PMSE by 1033.4. The PMSE of the selected isoprenoid genes in $\widehat{\boldsymbol{D}}$ is probably smaller than that of the full isoprenoid genes. We conclude that the multiple test procedure effectively works for selecting a set of significant genes.

Appendix A.

Throughout, we write that $\boldsymbol{x}_{0j}=\boldsymbol{x}_j-\boldsymbol{\mu}$, $x_{0j(*)}=x_{j(*)}-\mu_*$, $\boldsymbol{y}_{0j}=\sum_{i\neq i'}^r c_i\boldsymbol{\gamma}_{i'}w_{ij}w_{i'j}+c_*\sum_{i=1}^r\boldsymbol{\gamma}_iw_{ij}w_{j*}$ for each j, $\kappa=\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\Delta_L$ and $\boldsymbol{A}=\sigma_*^2\boldsymbol{\Sigma}+\boldsymbol{\sigma}\boldsymbol{\sigma}^T-2\sum_{i=1}^r c_i^2\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T$. Note that $\mathrm{Var}_{\boldsymbol{\theta}}(\boldsymbol{y}_{0j})=\boldsymbol{A}$ under (A-i) and (A-iii).

Lemma A.1. *Assume (A-i) to (A-iii). Then, we have as $p\to\infty$ and $n\to\infty$ that*

$$\begin{aligned}&Var_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})\\&=\Big(\frac{2\sigma_*^4tr(\boldsymbol{\Sigma}^2)}{n^2}+\frac{4}{n}\big\{\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma}+||\boldsymbol{\sigma}||^4+\sum_{i=1}^r(M_i-2)c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2\big\}\Big)\{1+o(1)\};\end{aligned}$$

$$\widehat{T}_{n,\boldsymbol{\sigma}}-||\boldsymbol{\sigma}||^2=\frac{2}{n(n-1)}\sum_{i<j}^n\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j}+o_p\big\{Var_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})^{1/2}\big\}\quad \textit{under (A-iv).}$$

Proof of Lemma A.1. We have that $\boldsymbol{x}_{0j}x_{0j(*)} - \boldsymbol{\sigma} = \boldsymbol{y}_{0j} + \sum_{i=1}^{r} c_i\boldsymbol{\gamma}_i(w_{ij}^2 - 1)$. Here, we note that $\sum_{i,j}^{r} c_i^2 c_j^2(\boldsymbol{\gamma}_i^T\boldsymbol{\gamma}_j)^2 \le \sigma_*^2\sum_{i=1}^{r} c_i^2\boldsymbol{\gamma}_i^T\boldsymbol{\Sigma}\boldsymbol{\gamma}_i \le \sigma_*^2(\sum_{i=1}^{r} c_i^4)^{1/2} \times(\sum_{i=1}^{r}\boldsymbol{\gamma}_i^T\boldsymbol{\Sigma}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\boldsymbol{\Sigma}\boldsymbol{\gamma}_i)^{1/2} \le \sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^4)^{1/2}$ from the facts that $\sum_{i=1}^{r} c_i^4 \le \sigma_*^4$ and $\sum_{i=1}^{r}\boldsymbol{\gamma}_i^T\boldsymbol{\Sigma}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T\boldsymbol{\Sigma}\boldsymbol{\gamma}_i \le \sum_{i=1}^{r}\boldsymbol{\gamma}_i^T\boldsymbol{\Sigma}^3\boldsymbol{\gamma}_i = \mathrm{tr}(\boldsymbol{\Sigma}^4)$. Then, it holds under (A-i) and (A-iii) that

$$\begin{aligned}&\mathrm{Var}_{\boldsymbol{\theta}}\big\{\textstyle\sum_{i=1}^{r} c_i\boldsymbol{\gamma}_i^T(w_{ij}^2-1)(\boldsymbol{x}_{0j'}x_{0j'(*)}-\boldsymbol{\sigma})\big\}\\ &\quad= O\{\textstyle\sum_{i=1}^{r} c_i^2\boldsymbol{\gamma}_i^T(\sigma_*^2\boldsymbol{\Sigma}+\boldsymbol{\sigma}\boldsymbol{\sigma}^T)\boldsymbol{\gamma}_i\} = o\{\sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^2)\} + O\{\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma}\}\end{aligned} \tag{A.1}$$

for $j \ne j'$. Let $\boldsymbol{y}_{1i(j)} = (x_{i(*)} - \overline{x}_{n(1*)(i+j)})(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n(1)(i+j)}) - n_{(1)}^{-1}(n_{(1)}-1)\boldsymbol{\sigma}$ and $\boldsymbol{y}_{2j(i)} = (x_{j(*)} - \overline{x}_{n(2*)(i+j)})(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(i+j)}) - n_{(2)}^{-1}(n_{(2)}-1)\boldsymbol{\sigma}$ for $i < j\ (\le n)$. Then, we can write that

$$\begin{aligned}\widehat{T}_{n,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2 =& 2u_n\{n(n-1)\}^{-1}\textstyle\sum_{i<j}^{n}\big[\boldsymbol{y}_{1i(j)}^T\boldsymbol{y}_{2j(i)}\\ &+ \boldsymbol{\sigma}^T\{\boldsymbol{y}_{1i(j)}(n_{(2)}-1)/n_{(2)} + \boldsymbol{y}_{2j(i)}(n_{(1)}-1)/n_{(1)}\}\big].\end{aligned} \tag{A.2}$$

Here, it holds from (A.1) that

$$\begin{aligned}&\mathrm{Var}_{\boldsymbol{\theta}}(\boldsymbol{y}_{1i(j)}^T\boldsymbol{y}_{2j(i)}) = \mathrm{tr}(\boldsymbol{A}^2)\{1+o(1)\}\\ &\qquad\qquad = \sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^2)\{1+o(1)\} + O(\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma} + ||\boldsymbol{\sigma}||^4);\\ &\mathrm{Var}_{\boldsymbol{\theta}}(u_n\boldsymbol{y}_{1i(j)}^T\boldsymbol{y}_{2j(i)} - \boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j}) = o\{\mathrm{tr}(\boldsymbol{A}^2)\};\\ &\mathrm{Var}_{\boldsymbol{\theta}}(\boldsymbol{\sigma}^T\boldsymbol{y}_{1i(j)}) = \{\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma} + ||\boldsymbol{\sigma}||^4 + \textstyle\sum_{i=1}^{r}(M_i-2)c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2\}\{1+o(1)\}\end{aligned}$$

for $i < j$. Then, from (A.1)-(A.2), we have that

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}}) =& \big[2n^{-2}\sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^2) + 4n^{-1}\{\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma}\\ &+ ||\boldsymbol{\sigma}||^4 + \textstyle\sum_{i=1}^{r}(M_i-2)c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2\}\big]\{1+o(1)\}\end{aligned}$$

under (A-i) to (A-iii), and

$$\widehat{T}_{n,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2 = 2\{n(n-1)\}^{-1}\textstyle\sum_{i<j}^{n}\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j} + o_p\big\{\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})^{1/2}\big\}$$

under (A-i) to (A-iv). It concludes the results. □

Lemma A.2. *Under (A-i), we have as $p \to \infty$ and $n \to \infty$ that $Var_{\boldsymbol{\theta}}(W_n) = [4n^{-2}tr(\boldsymbol{\Sigma}^2)^2 + 8n^{-1}\{tr(\boldsymbol{\Sigma}^4) + \sum_{i=1}^{r}(M_i-2)(\boldsymbol{\gamma}_i^T\boldsymbol{\Sigma}\boldsymbol{\gamma}_i)^2/2\}]\{1+o(1)\}$.*

Proof of Lemma A.2. By noting that $\sum_{i,j}^{r}(\boldsymbol{\gamma}_i^T\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^T\boldsymbol{\gamma}_i)^2 \le \sum_{i=1}^{r}(\boldsymbol{\gamma}_i^T\boldsymbol{\Sigma}\boldsymbol{\gamma}_i)^2 \le \mathrm{tr}(\boldsymbol{\Sigma}^4)$, we have that $\mathrm{Var}_{\boldsymbol{\theta}}\{\mathrm{tr}(\{(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})-\boldsymbol{\Sigma}\}\{(\boldsymbol{x}_j-\boldsymbol{\mu})(\boldsymbol{x}_j-\boldsymbol{\mu})-\boldsymbol{\Sigma}\})\} = 2\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 + O\{\mathrm{tr}(\boldsymbol{\Sigma}^4)\}$ $(i \ne j)$ and $\mathrm{Var}_{\boldsymbol{\theta}}\{\mathrm{tr}(\{(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})-\boldsymbol{\Sigma}\}\boldsymbol{\Sigma})\} = 2\mathrm{tr}(\boldsymbol{\Sigma}^4) + \sum_{i=1}^{r}(M_i-2)(\boldsymbol{\gamma}_i^T\boldsymbol{\Sigma}^2\boldsymbol{\gamma}_i)^2$. Thus in a way similar to the proof of Lemma A.1, we can conclude the result. $\square$

Lemma A.3. *Under (A-i) to (A-iv), we have as $p \to \infty$ and $n \to \infty$ that $E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j})^2(\boldsymbol{y}_{0i'}^T\boldsymbol{y}_{0j})^2\} = O\{tr(\sigma_*^4\boldsymbol{\Sigma}^2)^2\}$ for $i, i' \ne j$.*

Proof of Lemma A.3. Let $\boldsymbol{B} = \sigma_*^2\boldsymbol{\Sigma} + \boldsymbol{\sigma}\boldsymbol{\sigma}^T$. Let $c_{r+1} = c_*$ and $w_{r+1j} = w_{j*}$ for each j. We first consider the case when $i = i' \ne j$. Then, we write that

$$(\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j})^2 = \textstyle\sum_{k,k'}^{r}(\sum_{l=1(\ne k)}^{r+1} c_l w_{li})(\sum_{s=1(\ne k')}^{r+1} c_s w_{si})\boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_{k'}w_{ki}w_{k'i}.$$

Let $E(w_{ij}^3) = M_{3i}$ for $i = 1, ..., r+1$. Note that $|M_{3i}| \le \{E(w_{ij}^4)E(w_{ij}^2)\}^{1/2} < \infty$ under (A-i) and (A-iii) from Schwarz's inequality. Also note that

$$\textstyle\sum_{k,l}^{r}|c_kc_l\boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_l| \le \{\sum_{k,l}^{r}c_k^2c_l^2\}^{1/2}\{\sum_{k,l}^{r}(\boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j})^2(\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_l)^2\}^{1/2} \le \boldsymbol{y}_{0j}^T\boldsymbol{B}\boldsymbol{y}_{0j};$$

$$\textstyle\sqrt{\sigma_*^2}\sum_{k=1}^{r}|c_k\boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_{k'}| \le \{\boldsymbol{y}_{0j}^T\boldsymbol{B}\boldsymbol{y}_{0j}\sigma_*^2(\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_{k'})^2\}^{1/2} \le \boldsymbol{y}_{0j}^T\boldsymbol{B}\boldsymbol{y}_{0j};$$

$$|\boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_{k'}| \le \boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_k + \boldsymbol{\gamma}_{k'}^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_{k'}$$

w.p.1 and $\sum_{l=1}^{r+1}|c_l^3| \le (\sigma_*^2)^{3/2}$. Then, we can evaluate that

$$\begin{aligned}&\textstyle\sum_{l=1}^{r+1}\sum_{k=1(\ne l)}^{r}\sum_{k'=1(\ne k,l)}^{r}c_l^2c_{k'}^2|\boldsymbol{\gamma}_{k'}^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_kE(w_{kj}^3)E(w_{k'j}^3)|\\ &\le \kappa_1/2\textstyle\sum_{k=1}^{r}(\boldsymbol{y}_{0j}^T\boldsymbol{B}\boldsymbol{y}_{0j} + \sigma_*^2\boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_k)\sigma_*^2\boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_k \le \kappa_1(\boldsymbol{y}_{0j}^T\boldsymbol{B}\boldsymbol{y}_{0j})^2;\end{aligned}$$

$$\begin{aligned}&\textstyle\sum_{l=1}^{r+1}\sum_{k=1(\ne l)}^{r}\sum_{k'=1(\ne k,l)}^{r}|c_l^3c_k\boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_{k'}\boldsymbol{\gamma}_{k'}^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_{k'}E(w_{k'j}^3)E(w_{lj}^3)|\\ &\le \kappa_2\boldsymbol{y}_{0j}^T\boldsymbol{B}\boldsymbol{y}_{0j}\textstyle\sum_{k'=1}^{r}\sigma_*^2\boldsymbol{\gamma}_{k'}^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_{k'} \le \kappa_2(\boldsymbol{y}_{0j}^T\boldsymbol{B}\boldsymbol{y}_{0j})^2;\end{aligned}$$

$$\begin{aligned}&\textstyle\sum_{l=1}^{r}\sum_{k=1(\ne l)}^{r}\sum_{k'=1(\ne k,l)}^{r}|c_lc_k^2c_{k'}\boldsymbol{\gamma}_k^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_l\boldsymbol{\gamma}_{k'}^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_{k'}E(w_{kj}^3)E(w_{k'j}^3)|\\ &\le \kappa_3\boldsymbol{y}_{0j}^T\boldsymbol{B}\boldsymbol{y}_{0j}\textstyle\sum_{k'=1}^{r}\sigma_*^2\boldsymbol{\gamma}_{k'}^T\boldsymbol{y}_{0j}\boldsymbol{y}_{0j}^T\boldsymbol{\gamma}_{k'} \le \kappa_3(\boldsymbol{y}_{0j}^T\boldsymbol{B}\boldsymbol{y}_{0j})^2\end{aligned}$$

w.p.1 for some positive constants κ_1, κ_2 and κ_3. Then, it holds that

$$E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j})^4|\boldsymbol{y}_{0j}\} \le \kappa(\boldsymbol{y}_j^T\boldsymbol{B}\boldsymbol{y}_j)^2$$

w.p.1 for some positive constant κ. Hence, we have that

$$E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j})^4\} = E_{\boldsymbol{\theta}}[E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j})^4|\boldsymbol{y}_{0j}\}] \le E_{\boldsymbol{\theta}}\{\kappa(\boldsymbol{y}_j^T\boldsymbol{B}\boldsymbol{y}_j)^2\} = O\{\mathrm{tr}(\boldsymbol{B}^2)^2\}. \tag{A.3}$$

Next, we consider the case when $i \neq i' \neq j$. We have from (A.3) that $E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j})^2(\boldsymbol{y}_{0i'}^T\boldsymbol{y}_{0j})^2\} \le [E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j})^4\}E_{\boldsymbol{\theta}}\{(\boldsymbol{y}_{0i'}^T\boldsymbol{y}_{0j})^4\}]^{1/2} = O\{\mathrm{tr}(\boldsymbol{B}^2)^2\}$. Under (A-ii) and (A-iv), it holds that $\mathrm{tr}(\boldsymbol{B}^2)/\mathrm{tr}(\sigma_*^4\boldsymbol{\Sigma}^2) \to 1$ as $p \to \infty$. Thus it concludes the result. $\square$

Proof of Theorem 2.1. Let $y_{jn} = 2\{n(n-1)\}^{-1}\sum_{i=1}^{j-1}\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j}$ for $j = 2, ..., n$. Note that $\sum_{j=2}^{n} y_{jn} = 2\sum_{i<j}^{n}\boldsymbol{y}_{0i}^T\boldsymbol{y}_{0j}/\{n(n-1)\}$. Here, we have for $j = 3, ..., n$, that $E_{\boldsymbol{\theta}}(y_{jn}|y_{j-1n}, ..., y_{2n}) = 0$. Then, we consider applying the martingale central limit theorem given by McLeish [21]. Refer to Section 2.6 in Ghosh et al. [14] for the details of the martingale central limit theorem. Let $\delta = \sum_{j=2}^{n}\mathrm{Var}_{\boldsymbol{\theta}}(y_{jn}) = 2\mathrm{tr}(\boldsymbol{A}^2)/\{n(n-1)\}$. Then, it holds that $\mathrm{tr}(\boldsymbol{A}^2)/\{\sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^2)\} \to 1$ and $\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})/\delta \to 1$ as $p \to \infty$ and $n \to \infty$ under (A-i) to (A-iv). Let $v_j = y_{jn}/\delta^{1/2}$, $j = 2, ..., n$. Let $I(\cdot)$ denote the indicator function. It is necessary to check the following two conditions to apply the martingale central limit theorem:

- **(i)** $\sum_{j=2}^{n} E_{\boldsymbol{\theta}}\left\{v_j^2 I\left(v_j^2 > \tau\right)\right\} \to 0$ as $p \to \infty$ and $n \to \infty$ for any $\tau > 0$ (Lindeberg's condition);
- **(ii)** $\sum_{j=2}^{n} v_j^2 = 1 + o_p(1)$ as $p \to \infty$ and $n \to \infty$.

As for (i), note that $E_{\boldsymbol{\theta}}(v_j^4) = O\{(j-1)^2/n^4\}$ from Lemma A.3. Then, by using Chebyshev's inequality and Schwarz's inequality, for any $\tau > 0$, we have as $p \to \infty$ and $n \to \infty$ that

$$\begin{aligned}\sum_{j=2}^{n} E_{\boldsymbol{\theta}}\{v_j^2 I(v_j^2 > \tau)\} &\le \sum_{j=2}^{n}\left\{E_{\boldsymbol{\theta}}(v_j^4)P_{\boldsymbol{\theta}}\left(v_j^4 > \tau^2\right)\right\}^{1/2} \\ &\le \sum_{j=2}^{n} E_{\boldsymbol{\theta}}\left(v_j^4/\tau\right) = O\{\textstyle\sum_{j=2}^{n}(j-1)^2/n^4\} \to 0.\end{aligned}$$

As for (ii), note that

$$E_{\boldsymbol{\theta}}\Big[\Big\{v_i^2 - \frac{2(i-1)}{n(n-1)}\Big\}\Big\{v_j^2 - \frac{2(j-1)}{n(n-1)}\Big\}\Big] = O\Big\{\frac{(i-1)^2\mathrm{tr}(\boldsymbol{A}^4)}{n^4\mathrm{tr}(\boldsymbol{A}^2)^2} + \frac{(i-1)}{n^4}\Big\} \quad (A.4)$$

for $2 \le i < j \le n$. Note that $\mathrm{tr}(\boldsymbol{A}^4)/\mathrm{tr}(\sigma_*^2\boldsymbol{\Sigma}^2)^2 \to 0$ under (A-ii) and (A-iv). Then, from (A.4) and $E_{\boldsymbol{\theta}}(v_j^4) = O\{(j-1)^2/n^4\}$, we have as $p \to \infty$ and $n \to \infty$ that

$$\mathrm{Var}_{\boldsymbol{\theta}}(\textstyle\sum_{j=2}^{n} v_j^2) = O(n^{-1}) + O\{\mathrm{tr}(\boldsymbol{A}^4)/\mathrm{tr}(\boldsymbol{A}^2)^2\} \to 0$$

under (A-i) to (A-iv). Now, with the help of the martingale central limit theorem, under (A-i) to (A-iv), we have from Lemma A.1 that

$$\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})^{-1/2}(\widehat{T}_{n,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2) = \textstyle\sum_{j=2}^n v_j + o_p(1) \Rightarrow N(0,1)$$

as $p \to \infty$ and $n \to \infty$. It concludes the result. □

Proof of Corollary 2.1. We have under $\sigma_*^2 \mathrm{tr}(\boldsymbol{\Sigma}^2)^{1/2}/(n||\boldsymbol{\sigma}||^2) \to 0$ that

$$\frac{\sum_{i=1}^r c_i^2(\boldsymbol{\sigma}^T\boldsymbol{\gamma}_i)^2}{n||\boldsymbol{\sigma}||^4} \le \frac{\sigma_*^2 \boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma}}{n||\boldsymbol{\sigma}||^4} \le \frac{\sigma_*^2\lambda_1}{n||\boldsymbol{\sigma}||^2} \le \frac{\sigma_*^2 \mathrm{tr}(\boldsymbol{\Sigma}^2)^{1/2}}{n||\boldsymbol{\sigma}||^2} = o(1).$$

Thus by noting $\mathrm{tr}(\boldsymbol{\Sigma}^4)/\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 \le 1$, from the proof of Lemma A.1, we have that $\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}}/||\boldsymbol{\sigma}||^2) \to 0$ under (A-i) and (A-iii) without (A-ii). Hence, from Chebyshev's inequality, it concludes the results. □

Proof of Corollary 2.2. We have as $p \to \infty$ and $n \to \infty$ that

$$\mathrm{Var}_{\boldsymbol{\theta}}\Big(\frac{W_n}{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\Big) = O(n^{-2}) + O\Big(\frac{\mathrm{tr}(\boldsymbol{\Sigma}^4)}{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 n}\Big) \to 0 \quad \text{and} \quad \mathrm{Var}_{\boldsymbol{\theta}}(S_{n(*)}) \to 0.$$

Then, from Chebyshev's inequality, it holds that $W_n/\mathrm{tr}(\boldsymbol{\Sigma}^2) = 1 + o_p(1)$ and $S_{n(*)} = \sigma_*^2 + o_p(1)$. Hence, from Theorem 2.1, we obtain that

$$\frac{\widehat{T}_{n,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2}{S_{n(*)}\sqrt{2W_n}/n} = \frac{\widehat{T}_{n,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2}{\sigma_*^2\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}/n} + o_p(1) \Rightarrow N(0,1).$$

It concludes the result. □

Proof of Theorem 2.2. We first consider the case when $\boldsymbol{\rho} = \mathbf{0}$. From Corollary 2.2, we have that size$= P_{\boldsymbol{\theta}}\{\widehat{T}_{n,\boldsymbol{\sigma}}/(S_{n(*)}\sqrt{2W_n}/n) > z_\alpha\} = \alpha + o(1)$. Next, we consider the case when $\boldsymbol{\rho} \ne \mathbf{0}$. From Corollary 2.2, we have that

$$\begin{aligned}\text{power} &= P_{\boldsymbol{\theta}}\big\{N(0,1) > z_\alpha - (S_{n(*)}\sqrt{2W_n})^{-1} n||\boldsymbol{\sigma}||^2\big\} + o(1) \\ &= \Phi\big\{(\sigma_*^2\{2\mathrm{tr}(\boldsymbol{\Sigma}^2)\}^{1/2})^{-1} n||\boldsymbol{\sigma}||^2 - z_\alpha\big\} + o(1).\end{aligned}$$

It concludes the results. □

Proof of Corollary 2.3. From Corollary 2.1, we have that

$$\text{power} = P_{\boldsymbol{\theta}}\Big(\frac{\widehat{T}_{n,\boldsymbol{\sigma}}}{||\boldsymbol{\sigma}||^2} > \frac{z_\alpha\sigma_*^2\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}}{n||\boldsymbol{\sigma}||^2}\{1 + o_p(1)\}\Big) = P_{\boldsymbol{\theta}}\{1 > o_p(1)\} \to 1.$$

It concludes the results. □

Proof of Theorem 3.1. We first consider the case when $\boldsymbol{\rho} = \mathbf{0}$. We have from (12) that $(\sigma_*^2\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}/n)/\{\kappa/(z_\alpha + z_\beta)\} \le 1$. Then, from the facts that $S_{n(*)} = \sigma_*^2 + o_p(1)$ and $\mathrm{tr}(\boldsymbol{S}_n)/\mathrm{tr}(\boldsymbol{\Sigma}) = 1 + o_p(1)$, it holds as $p \to \infty$ that

$$\text{size} = P_{\boldsymbol{\theta}}\Big(\widehat{T}_{n,\boldsymbol{\sigma}} > \frac{S_{n(*)}\mathrm{tr}(\boldsymbol{S}_n)\Delta_L z_\alpha}{z_\alpha + z_\beta}\Big) \le P_{\boldsymbol{\theta}}\big(N(0,1) > z_\alpha\big) + o(1) \to \alpha$$

by using Theorem 2.1. Next, we consider the case when $||\boldsymbol{\sigma}||^2 \ge \kappa$. By noting that (A-iv) holds when $\liminf_{p\to\infty} \kappa/||\boldsymbol{\sigma}||^2 > 0$ and $\liminf C/n > 0$, from Theorem 2.1, it holds as $p \to \infty$ that

$$\begin{aligned}
\text{power} &= P_{\boldsymbol{\theta}}\Big(\widehat{T}_{n,\boldsymbol{\sigma}} > \frac{S_{n(*)}\mathrm{tr}(\boldsymbol{S}_n)\Delta_L z_\alpha}{z_\alpha + z_\beta}\Big) \\
&= P_{\boldsymbol{\theta}}\Big(\frac{\widehat{T}_{n,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2}{\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})^{1/2}} > \frac{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\{1 + o_p(1)\}\Delta_L z_\alpha - (z_\alpha + z_\beta)||\boldsymbol{\sigma}||^2}{\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}})^{1/2}(z_\alpha + z_\beta)}\Big) \\
&\ge P_{\boldsymbol{\theta}}\big(N(0,1) > -z_\beta\big) + o(1) \to 1 - \beta
\end{aligned}$$

when $\liminf C/n > 0$ and $\liminf \kappa/||\boldsymbol{\sigma}||^2 > 0$. When $C/n \to 0$ or $\kappa/||\boldsymbol{\sigma}||^2 \to 0$ as $p \to \infty$, from the fact that

$$\mathrm{Var}_{\boldsymbol{\theta}}(\widehat{T}_{n,\boldsymbol{\sigma}}) = O\{(\kappa C/n)^2\} + O\{||\boldsymbol{\sigma}||^2(\kappa C/n)\} + o(||\boldsymbol{\sigma}||^4) = o(||\boldsymbol{\sigma}||^4),$$

it holds as $p \to \infty$ that

$$\begin{aligned}
\text{power} = P_{\boldsymbol{\theta}}\Big(\widehat{T}_{n,\boldsymbol{\sigma}} > \frac{S_{n(*)}\mathrm{tr}(\boldsymbol{S}_n)\Delta_L z_\alpha}{z_\alpha + z_\beta}\Big) &= P_{\boldsymbol{\theta}}\Big(\frac{\widehat{T}_{n,\boldsymbol{\sigma}}}{||\boldsymbol{\sigma}||^2} > \frac{S_{n(*)}\mathrm{tr}(\boldsymbol{S}_n)\Delta_L z_\alpha}{(z_\alpha + z_\beta)||\boldsymbol{\sigma}||^2}\Big) \\
&\ge P_{\boldsymbol{\theta}}\Big(1 + o_p(1) > \frac{z_\alpha}{z_\alpha + z_\beta}\Big) \to 1.
\end{aligned}$$

Thus we have that power $\ge 1 - \beta + o(1)$ when $||\boldsymbol{\sigma}||^2/\{\sigma_*^2\mathrm{tr}(\boldsymbol{\Sigma})\} \ge \Delta_L$. It concludes the results. □

Lemma A.4. *Let*

$$\widetilde{T}_{N,\boldsymbol{\sigma}} = \frac{2}{N(N-1)}\sum_{i<j}^{N} x_{0i(*)}\boldsymbol{x}_{0i}^T x_{0j(*)}\boldsymbol{x}_{0j}.$$

Assume (A-i) to (A-iii). Assume also that $\limsup_{p\to\infty} ||\boldsymbol{\sigma}||^2/\kappa < \infty$. *For the two-stage procedure given by (16)-(17), it holds as* $p \to \infty$ *that*

$$\frac{\widetilde{T}_{N,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2}{\sigma_*^2\sqrt{2tr(\boldsymbol{\Sigma}^2)/N}} \Rightarrow N(0,1).$$

Proof of Lemma A.4. By using Chebyshev's inequality, from Lemma A.2, for any $\tau > 0$, it holds as $p \to \infty$ that $P_{\boldsymbol{\theta}}\{|W_m - \mathrm{tr}(\boldsymbol{\Sigma}^2)| > \tau \mathrm{tr}(\boldsymbol{\Sigma}^2)/C_\star^{1/2}\} \to 0$ under (A-i)-(A-ii) and (16). Thus we have $W_m = \mathrm{tr}(\boldsymbol{\Sigma}^2) + o_p\{\mathrm{tr}(\boldsymbol{\Sigma}^2)/C_\star^{1/2}\}$. Here, by using Hölder's inequality, we also have that $\mathrm{tr}(\boldsymbol{\Sigma}^2) = \sum_{i=1}^p \lambda_i^2 \le (\sum_{i=1}^p \lambda_i^4)^{1/3}(\sum_{i=1}^p \lambda_i)^{2/3}$. Then it holds that

$$\mathrm{tr}(\boldsymbol{\Sigma}^2)/\mathrm{tr}(\boldsymbol{\Sigma})^2 \le \mathrm{tr}(\boldsymbol{\Sigma}^4)/\mathrm{tr}(\boldsymbol{\Sigma}^2)^2. \tag{A.5}$$

Thus it follows under (A-i)-(A-ii) and (16) that

$$P_{\boldsymbol{\theta}}\{|\mathrm{tr}(\boldsymbol{S}_m) - \mathrm{tr}(\boldsymbol{\Sigma})| > \tau \mathrm{tr}(\boldsymbol{\Sigma})/C_\star^{1/2}\} = O\{(C_\star/m)\mathrm{tr}(\boldsymbol{\Sigma}^2)/\mathrm{tr}(\boldsymbol{\Sigma})^2\} \to 0.$$

Then, we have $\mathrm{tr}(\boldsymbol{S}_m) = \mathrm{tr}(\boldsymbol{\Sigma}) + o_p\{\mathrm{tr}(\boldsymbol{\Sigma})/C_\star^{1/2}\}$. Let $Y = (z_\alpha + z_\beta)\sqrt{2W_m} /\{\Delta_L \mathrm{tr}(\boldsymbol{S}_m)\} + 2\eta(z_\alpha + z_\beta)^2$. Then, by noting that $Y/C_\star = 1 + o_p(C_\star^{-1/2})$ and $m/C_\star \le 1$, we have that $|N - C_\star| = o_p(C_\star^{1/2})$. Then, we write that $|N - C_\star| = O_p(\omega C_\star^{1/2})$, where ω is a variable such that $\omega \to 0$ as $p \to \infty$. Let $C_L = \lfloor C_\star - (\omega C_\star)^{1/2} \rfloor$ and $C_U = \lceil C_\star + (\omega C_\star)^{1/2} \rceil$. It holds as $p \to \infty$ that $C_L < N < C_U$ w.p.1. Let $\boldsymbol{u}_j = x_{0j(*)}\boldsymbol{x}_{0j}$. Now, we write that

$$\widetilde{T}_{N,\boldsymbol{\sigma}} = \sum_{i<j}^{C_L} \frac{2\boldsymbol{u}_i^T \boldsymbol{u}_j}{N(N-1)} + \sum_{j=C_L+1}^{N} \sum_{i=1}^{C_L} \frac{2\boldsymbol{u}_i^T \boldsymbol{u}_j}{N(N-1)} + \sum_{i \ne j (>C_L)}^{N} \frac{\boldsymbol{u}_i^T \boldsymbol{u}_j}{N(N-1)}. \tag{A.6}$$

Note that $\boldsymbol{\sigma}^T \boldsymbol{\Sigma} \boldsymbol{\sigma} \le ||\boldsymbol{\sigma}||^2 \lambda_1 \le ||\boldsymbol{\sigma}||^2 \mathrm{tr}(\boldsymbol{\Sigma}^2)^{1/2}$. Then, by using Chebyshev's inequality and Schwarz's inequality, for any $\tau > 0$, we have that

$$\begin{aligned} &P_{\boldsymbol{\theta}}\big(\big|\textstyle\sum_{j=C_L+1}^{N}\sum_{i=1}^{C_L}(\boldsymbol{u}_i^T\boldsymbol{u}_j - ||\boldsymbol{\sigma}||^2)/C_\star^2\big| > \tau\kappa\big) \\ &\le P_{\boldsymbol{\theta}}\big(\textstyle\sum_{j=C_L+1}^{C_U}\big|\sum_{i=1}^{C_L}(\boldsymbol{u}_i^T\boldsymbol{u}_j - ||\boldsymbol{\sigma}||^2)/C_\star^2\big| > \tau\kappa\big) + o(1) \\ &= O\big\{\omega\sigma_*^4\mathrm{tr}(\boldsymbol{\Sigma}^2)/(C_\star^2\kappa^2)\big\} + O\big\{\omega(\sigma_*^2\boldsymbol{\sigma}^T\boldsymbol{\Sigma}\boldsymbol{\sigma} + ||\boldsymbol{\sigma}||^4)/(C_\star^2\kappa^2)\big\} + o(1) \to 0 \end{aligned} \tag{A.7}$$

under (A-i) to (A-iii) and $\limsup_{p\to\infty} ||\boldsymbol{\sigma}||^2/\kappa < \infty$. Thus by noting that $N/C_\star = 1 + o_p(1)$, we obtain that

$$\textstyle\sum_{j=C_L+1}^{N}\sum_{i=1}^{C_L}(\boldsymbol{u}_i^T\boldsymbol{u}_j - ||\boldsymbol{\sigma}||^2)/\{N(N-1)\} = o_p(\kappa).$$

Similarly to (A.7), for any $\tau > 0$, we obtain that $P_{\boldsymbol{\theta}}\{|\sum_{i\ne j(>C_L)}^{N}(\boldsymbol{u}_i^T\boldsymbol{u}_j - ||\boldsymbol{\sigma}||^2)/C_\star^2| > \tau\kappa\} \to 0$. Hence, we have that $\sum_{i\ne j(>C_L)}^{N}(\boldsymbol{u}_i^T\boldsymbol{u}_j - ||\boldsymbol{\sigma}||^2)/\{N(N-$

$1)\} = o_p(\kappa)$. Note that $N/C_L = 1 + o_p(1)$. Then, from (A.6), it holds as $p \to \infty$ that $\widetilde{T}_{N,\boldsymbol{\sigma}} = \widetilde{T}_{C_L,\boldsymbol{\sigma}} + o_p(\kappa)$. Therefore, similarly to the proof of Theorems 2.1 and 3.1, under (A-i) to (A-iii), we have that

$$\frac{\widetilde{T}_{N,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2}{\sigma_*^2\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)/N}} = \frac{\widetilde{T}_{C_L,\boldsymbol{\sigma}} - ||\boldsymbol{\sigma}||^2}{\sigma_*^2\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)/C_L}} + o_p(1) \Rightarrow N(0,1).$$

It concludes the result. $\square$

Lemma A.5. *Assume (A-i) to (A-iii). Assume also that* $\limsup_{p\to\infty} ||\boldsymbol{\sigma}||^2 /\kappa < \infty$. *For the two-stage procedure given by (16)-(17), it holds as* $p \to \infty$ *that*

$$\widehat{T}_{N,\boldsymbol{\sigma}} = \widetilde{T}_{N,\boldsymbol{\sigma}} + o_p(\kappa),$$

where $\widetilde{T}_{N,\boldsymbol{\sigma}}$ *is the one given in Lemma A.4.*

Proof of Lemma A.5. Let $k_\star = \lceil k/2 - 1\rceil$. We write that

$$\widehat{T}_{N,\boldsymbol{\sigma}} = \frac{2u_N}{N(N-1)} \sum_{k=3}^{2N-1} \sum_{i=\max\{1,k-N\}}^{k_\star} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{N(1)(k)})^T(\boldsymbol{x}_{k-i} - \overline{\boldsymbol{x}}_{N(2)(k)})$$
$$\times (x_{i(*)} - \overline{x}_{N(1*)(k)})(x_{k-i(*)} - \overline{x}_{N(2*)(k)}).$$

Let $\boldsymbol{W}_{N(l)(k)1} = \boldsymbol{V}_{C_L(l)(k')} \setminus (\boldsymbol{V}_{N(l)(k)} \cap \boldsymbol{V}_{C_L(l)(k')})$ and $\boldsymbol{W}_{N(l)(k)2} = \boldsymbol{V}_{N(l)(k)} \setminus (\boldsymbol{V}_{N(l)(k)} \cap \boldsymbol{V}_{C_L(l)(k')})$, where $k' = \min\{k, 2C_L - 1\}$ and C_L is defined in the proof of Lemma A.4. Note that $C_L < N < C_U$ w.p.1 as $p \to \infty$, where C_U is defined in the proof of Lemma A.4. Then, it holds for $j = 1,2$, that $\#(\boldsymbol{W}_{N(l)(k)j}) \le C_U - C_L = o(C_\star^{1/2})$ w.p.1 as $p \to \infty$. Now, we write that

$$\overline{\boldsymbol{x}}_{N(l)(k)} = \sum_{j\in\boldsymbol{V}_{C_L(l)(k')}} \boldsymbol{x}_j/N_{(l)} - \sum_{j\in\boldsymbol{W}_{N(l)(k)1}} \boldsymbol{x}_j/N_{(l)} + \sum_{j\in\boldsymbol{W}_{N(l)(k)2}} \boldsymbol{x}_j/N_{(l)}.$$

Let $C_{L(1)} = \lceil C_L/2\rceil$. Let $\psi_{ijk} = C_{L(1)}^{-1}(\boldsymbol{x}_j - \boldsymbol{\mu})^T(\boldsymbol{x}_{k-i} - \boldsymbol{\mu})(x_{i(*)} - \mu_*)(x_{k-i(*)} - \mu_*)$. Then, in a way similar to (A.7), for any $\tau > 0$, we have for $i' = 1,2$, that

$$P_{\boldsymbol{\theta}}\Big(C_L^{-2}\Big|\sum_{k=3}^{2N-1}\sum_{i=\max\{1,k-N\}}^{k_\star}\sum_{j\in\boldsymbol{W}_{N(1)(k)i'}}\psi_{ijk}\Big|>\tau\kappa\Big)$$

$$\leq P_{\boldsymbol{\theta}}\Big\{C_L^{-2}\sum_{k=3}^{2C_L-1}\sum_{j\in\boldsymbol{W}_{N(1)(k)i'}}\Big(\Big|\sum_{i=\max\{1,k-C_L\}}^{k_\star}\psi_{ijk}\Big|+\sum_{i=\max\{1,k-N\}}^{\max\{1,k-C_L\}-1}|\psi_{ijk}|\Big)>$$

$$\frac{\tau\kappa}{2}\Big\}+P_{\boldsymbol{\theta}}\Big(C_L^{-2}\sum_{k=2C_L}^{2N-1}\sum_{j\in\boldsymbol{W}_{N(1)(k)i'}}\sum_{i=\max\{1,k-N\}}^{k_\star}|\psi_{ijk}|>\frac{\tau\kappa}{2}\Big)\to 0.$$

Similarly, we have that

$$P_{\boldsymbol{\theta}}\Big(C_L^{-2}\Big|\sum_{k=3}^{2N-1}\sum_{i=\max\{1,k-N\}}^{k_\star}\sum_{j\in\boldsymbol{V}_{C_L(1)(k')}}\psi_{ijk}\Big|>\tau\kappa\Big)\to 0.$$

Then, we obtain that $\sum_{k=3}^{2N-1}\sum_{i=\max\{1,k-N\}}^{k_\star}(\overline{\boldsymbol{x}}_{N(1)(k)}-\boldsymbol{\mu})^T(\boldsymbol{x}_{k-i}-\boldsymbol{\mu})(x_{i(*)}-\mu_*)(x_{k-i(*)}-\mu_*)/\{N(N-1)\}=o_p(\kappa)$. Similarly, we obtain that

$$\begin{aligned}\widehat{T}_{N,\boldsymbol{\sigma}}&=\frac{2u_N}{N(N-1)}\sum_{k=3}^{2N-1}\sum_{i=\max\{1,k-N\}}^{k_\star}(\boldsymbol{x}_i-\overline{\boldsymbol{x}}_{N(1)(k)})^T(\boldsymbol{x}_{k-i}-\overline{\boldsymbol{x}}_{N(2)(k)})\\&\qquad\times(x_{i(*)}-\overline{x}_{N(1*)(k)})(x_{k-i(*)}-\overline{x}_{N(2*)(k)})\\&=2\sum_{k=3}^{2N-1}\sum_{i=\max\{1,k-N\}}^{k_\star}\frac{(x_{i(*)}-\mu_*)(\boldsymbol{x}_i-\boldsymbol{\mu})^T(\boldsymbol{x}_{k-i}-\boldsymbol{\mu})(x_{k-i(*)}-\mu_*)}{N(N-1)}+o_p(\kappa)\\&=\widetilde{T}_{N,\boldsymbol{\sigma}}+o_p(\kappa).\end{aligned}$$

It concludes the result. □

Remark 15. Assume (A-i) to (A-iii) and $\kappa/||\boldsymbol{\sigma}||^2\to 0$ as $p\to\infty$. For the two-stage procedure given by (16)-(17), it holds as $p\to\infty$ and $n\to\infty$ that $\widehat{T}_{N,\boldsymbol{\sigma}}/||\boldsymbol{\sigma}||^2=1+o_p(1)$.

Proof of Theorem 3.2. When $\limsup_{p\to\infty}||\boldsymbol{\sigma}||^2/\kappa<\infty$, by combining Lemmas A.4 and A.5, we have that $(\widehat{T}_{N,\boldsymbol{\sigma}}-||\boldsymbol{\sigma}||^2)/\{\sigma_*^2\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}/N\}\Rightarrow N(0,1)$. In a way similar to (A.7), it holds that $S_{N(*)}=\sigma_*^2+o_p(1)$ and $\mathrm{tr}(\boldsymbol{S}_N)/\mathrm{tr}(\boldsymbol{\Sigma})=$

$1+o_p(1)$. Note that $(1-2\eta(z_\alpha+z_\beta)^2/N)=1+o_p(1)$. Then, similarly to the proof of Theorem 3.1, the result is obtained. When $\kappa/||\boldsymbol{\sigma}||^2 \to 0$ as $p \to \infty$, with the help of Remark 15, it concludes the result. $\square$

Proof of Theorem 4.1. We first consider the subscript i having $\boldsymbol{\rho}_i = \mathbf{0}$. We have from (23) that $(\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}/N)/\{\mathrm{tr}(\boldsymbol{\Sigma})\Delta_L/(z_{\alpha/K}+z_{\beta/K})\} = 1+o_p(1)$. From Lemmas A.4 and A.5, under (A-i), (A-ii) and (A-v), we have that

$$P_{\boldsymbol{\theta}_K}\big(\widehat{T}_{N,\boldsymbol{\sigma}(i)} > \frac{\mathrm{tr}(\boldsymbol{S}_N)S_{iN(*)}\Delta_L z_{\alpha/K}}{z_{\alpha/K}+z_{\beta/K}}\big) \le P_{\boldsymbol{\theta}_K}\{N(0,1) > z_{\alpha/K}\} + o(1) \to \alpha/K$$

from the facts that $S_{iN(*)} = \sigma_{i*}^2 + o_p(1)$ and $\mathrm{tr}(\boldsymbol{S}_N)/\mathrm{tr}(\boldsymbol{\Sigma}) = 1+o_p(1)$. Then, by using Bonferroni's inequality, we obtain that

$$P_{\boldsymbol{\theta}_K}(\boldsymbol{D}^c \cap \widehat{\boldsymbol{D}} = \emptyset) \ge 1-\alpha+o(1).$$

Thus we have that $P_{\boldsymbol{\theta}_K}(\boldsymbol{D}^c \cap \widehat{\boldsymbol{D}} \neq \emptyset) \le \alpha + o(1)$.

Next, we consider the subscript i having $\boldsymbol{\rho}_i \neq \mathbf{0}$. Similarly to the proofs of Theorems 3.1 and 3.2, under (A-i), (A-ii) and (A-v), we have that

$$P_{\boldsymbol{\theta}_K}\big(\widehat{T}_{N,\boldsymbol{\sigma}(i)} > \frac{\mathrm{tr}(\boldsymbol{S}_N)S_{iN(*)}\Delta_L z_{\alpha/K}}{z_{\alpha/K}+z_{\beta/K}}\big) \ge 1-\beta/K+o(1) \text{ when } \frac{||\boldsymbol{\sigma}_i||^2}{\sigma_{i*}^2\mathrm{tr}(\boldsymbol{\Sigma})} \ge \Delta_L.$$

Thus by using Bonferroni's inequality, we obtain that $P_{\boldsymbol{\theta}_K}(\boldsymbol{D} \subseteq \widehat{\boldsymbol{D}}) \ge 1-\beta+o(1)$ when $\min_{i\in\boldsymbol{D}} ||\boldsymbol{\sigma}_i||^2/\{\sigma_{i*}^2\mathrm{tr}(\boldsymbol{\Sigma})\} \ge \Delta_L$. It concludes the results. $\square$

Acknowledgment

The authors would like to thank the anonymous referees for their constructive comments. Research of the first author was partially supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 23740066. Research of the second author was partially supported by Grants-in-Aid for Scientific Research (B) and Challenging Exploratory Research, JSPS, under Contract Numbers 22300094 and 23650142.

References

[1] J. Ahn, J.S. Marron, K.M. Muller, Y.-Y. Chi, The high-dimension, low-sample-size geometric representation holds under mild conditions, Biometrika 94 (2007) 760-766.

[2] M. Aoshima, K. Yata, Two-stage procedures for high-dimensional data, Sequential Anal. (*Editor's special invited paper*) 30 (2011) 356-399.

[3] M. Aoshima, K. Yata, Authors' response, Sequential Anal. 30 (2011) 432-440.

[4] M. Aoshima, K. Yata, Effective methodologies for statistical inference on microarray studies, in: P.E. Spiess (Ed.), Prostate Cancer - From Bench to Bedside, InTech, 2011, pp. 13-32.

[5] Z. Bai, H. Saranadasa, Effect of high dimension: By an example of a two sample problem, Statist. Sinica 6 (1996) 311-329.

[6] J. Baik, J.W. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, J. Multivariate Anal. 97 (2006) 1382-1408.

[7] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Roy. Statist. Soc. Ser. B 57 (1995) 289-300.

[8] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, Ann. Statist. 29 (2001) 1165-1188.

[9] S.X. Chen, Y.-L. Qin, A two-sample test for high-dimensional data with applications to gene-set testing, Ann. Statist. 38 (2010) 808-835.

[10] S. Chiaretti, X. Li, R. Gentleman, A. Vitale, M. Vignetti, F. Mandelli, J. Ritz, R. Foa, Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival, Blood 103 (2004) 2771-2778.

[11] M. Drton, M.D. Perlman, Multiple testing and error control in Gaussian graphical model selection, Statist. Sci. 22 (2007) 430-449.

[12] B. Efron, Better Bootstrap Confidence Intervals, J. Amer. Statist. Assoc. 82 (1987) 171-185.

[13] Y. Fujikoshi, V. Ulyanov, R. Shimizu, Multivariate Statistics: High-Dimensional and Large-Sample Approximations, Wiley, Hoboken, New Jersey, 2010.

[14] M. Ghosh, N. Mukhopadhyay, P.K. Sen, Sequential Estimation, Wiley, New York, 1997.

[15] N.G. Gottardo, K. Hoffmann, A.H. Beesley, J.R. Freitas, M.J. Firth, K.U. Perera, N.H. de Klerk, D.L. Baker, U.R. Kees, Identification of novel molecular prognostic markers for paediatric T-cell acute lymphoblastic leukaemia, Br. J. Haematol. 137 (2007) 319-328.

[16] P. Hall, J.S. Marron, A. Neeman, Geometric representation of high dimension, low sample size data, J. R. Statist. Soc. Ser. B 67 (2005) 427-444.

[17] A. Hero, B. Rajaratnam, Large-scale correlation screening, J. Amer. Statist. Assoc. 106 (2011) 1540-1552.

[18] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Statist. 6 (1979) 65-70.

[19] I. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, Ann. Statist. 29 (2001) 295-327.

[20] S. Jung, J.S. Marron, PCA consistency in high dimension, low sample size context, Ann. Statist. 37 (2009) 4104-4130.

[21] D.L. McLeish, Dependent central limit theorems and invariance principles, Ann. Probab. 2 (1974) 620-628.

[22] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, Statist. Sinica 17 (2007) 1617-1642.

[23] K.S. Pollard, S. Dudoit, M.J. van der Laan, Multiple testing procedures: R multtest package and applications to genomics, in: R. Gentleman, V. Carey, W. Huber, R. Irizarry, S. Dudoit (Eds.), Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Springer, New York, 2005, pp. 249-271.

[24] M.S. Srivastava, Some tests concerning the covariance matrix in high dimensional data, J. Japan Statist. Soc. 35 (2005) 251-272.

[25] M.S. Srivastava, Multivariate theory for analyzing high dimensional data, J. Japan Statist. Soc. 37 (2007) 53-86.

[26] G.J. Székely, M.L. Rizzo, Brownian distance covariance, Ann. Appl. Probab. 3 (2009) 1236-1265.

[27] A. Wille, P. Zimmermann, E. Vranova, A. Fürholz, O. Laule, S. Bleuler, L. Hennig, A. Prelic, P. von Rohr, L. Thiele, E. Zitzler, W. Gruissem, P. Bühlmann, Sparse graphical Gaussian modeling of the isoprenoid gene network in Arabidopsis thaliana, Genome Biol. 5 (2004) R92.

[28] K. Yata, Effective two-stage estimation for a linear function of high-dimensional Gaussian means, Sequential Anal. 29 (2010) 463-482.

[29] K. Yata, M. Aoshima, PCA consistency for non-Gaussian data in high dimension, low sample size context, Commun. Statist. Theory Methods, Special Issue Honoring S. Zacks (Ed. N. Mukhopadhyay) 38 (2009) 2634-2652.

[30] K. Yata, M. Aoshima, Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix, J. Multivariate Anal. 101 (2010) 2060-2077.

[31] K. Yata, M. Aoshima, Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations, J. Multivariate Anal. 105 (2012) 193-215.

[32] P.S. Zhong, S.X. Chen, Tests for high-dimensional regression coefficients with factorial designs, J. Amer. Statist. Assoc. 106 (2011) 260-274.